

Journal

Annual Open Access Publication of the School of Doctoral Studies of the European Union, Brussels, Belgium; Published by the IIU Press and Research Centre, A.C.

	Editorial Note	1
	Instructions for Contributors	2
Business Management and Economics Page 5	Empirical Evaluation of the Evidence of the Beneficial Influence of the Strategic Planning Process on the Overall Performance of Emerging Countries Companies <i>Walter Gerard Amedzro St-Hilaire</i>	7
	Institutional Infrastructure and the Adoption of International Financial Reporting Standards (IFRS) in Nigeria <i>Iyoha, F.O., Jafaru Jimoh</i>	17
	Comparing Bankruptcy prediction models in Iran <i>Ali Ebrahimi Kordlar, Nader Nikbakht</i>	24
	The Relationship Between Service Quality and Satisfaction on Customer Loyalty in Malaysian Mobile Communication Industry <i>Ahmed Audu Maiyaki, Norzaini Bt Mohd Noor, Sany Sanuri Mohd Mokhtar</i>	32
	Electronic Commerce, Automation and Online Banking in Nigeria: Challenges and Benefits <i>Stephen Ojeka, Ailemen Ikpefan O.</i>	39
	A Survey on the Assessment of Iranian External Auditors' Recognition of Materiality Levels <i>Azar Rahdarian, Hamed Hamedian</i>	51
	Small Scale Supplier Satisfaction: An Explorative Finding from Indian Manufacturing Industry <i>Manoj Kumar Mohanty, P. Gahan</i>	65
Engineering and Technology Page 77	Globalization and Energy Demands in the 21st Century <i>Andrew R. Thorp</i>	78
	Overview and Analysis on Cyber Terrorism <i>Steve Saint-Claire</i>	85
	The Impact of Computer Aided Design on Human Factors in Aviation Safety <i>Anne Curtney</i>	99
Science Page 111	Enhanced Wellbeing amongst Engineering Students through Nadi Shodhan Pranayama (Alternate Nostril Breathing) Training : An Analysis <i>Anurag Joshi, Mandeep Singh, Bharat Bhushan Singla, Sunil Joshi</i>	112
	An Effective Neuroprotective Treatment in Ischemic Stroke and Cerebral Trauma with Low Doses of L-Arginine, Lamotrigine and Tianeptine <i>Héctor Jara, Francisco García, Franky Torres, Fuad Lechin</i>	121
	Metastatic Breast Cancer: An evaluation of the Safety and Effectiveness of High Dose Chemotherapy with Autologous Stem Cell Transplant <i>Martin F. Taylor</i>	134
	Natural Gas Sweetening Process Design <i>Kurt Mearkeltor</i>	144
Social Science Page 157	Foreign Aid and Economic Development <i>Lawrence McMillan</i>	158
	Marcel Duchamp Artist's Works Profile and Analysis on the Nude in Art proposed by Duchamp and Cocteau <i>Erika González-Ehrlich</i>	166
	The Effect of Islamic Azad University Activity on Branch of Trade Case Study: Firoozabad <i>Mehrnoush Saadat Abadi, Nima Naddaf, Fateme Saeed Firoozabadi</i>	179
	Service Failure and Recovery: Comparison Between Health Care and Automobile Service Station <i>Anupam Krishna, G.S. Dangayach, Rakesh Jain</i>	184

FOR PRINTED VERSION: ISSN 1918-8722

FOR ONLINE VERSION: ISSN 1918-8730



School of Doctoral Studies (European Union) Journal is licensed under a Creative Commons Attribution 2.0 UK: England & Wales License. Further tips for using the supplied HTML and RDF are here:
<http://creativecommons.org/learn/technology/usingmarkup>

Editorial Note

SDS

School of Doctoral Studies
European Union



euasc

eu analogue standards
certification committee



Seal of the Scholarly Publishing
and Academic Resources Coalition
(Granted to the Business Intelligence
Journal on the 20th day of August, 2008)



DOAJ

DIRECTORY OF
OPEN ACCESS
JOURNALS

Lund University Libraries Sweden.

Sayco

Business Intelligence Service
London, UK

School of Doctoral Studies (European Union) Journal (SDSJ) publishes research analysis and inquiry into issues of importance to the academic and scientific community. Articles in SDSJ examine emerging trends and concerns in the areas of business management, economics, engineering, technology, natural and social science. The goal of SDSJ is to broaden the knowledge of students, academicians and society in general by promoting free access and provide valuable insight to human knowledge related information, research and ideas, presenting what the School of Doctoral Studies of the European Union's Faculty Members may consider as relevant contributions to human knowledge. SDSJ is an annual publication and all articles are peer-reviewed. SDSJ will be published annually (one volume per year) by the IIU Press and Research Centre, A.C. for the School of Doctoral Studies of the European Union, hosted at the Isles Internationales Université (European Union) in Brussels, Belgium.

School of Doctoral Studies of the European Union

Brussels EU Parliament Building

Square de Meeus

37 - 4th Floor

1000 Brussels, Belgium

<http://www.iiuedu.eu/school/about.html>

IIU Press and Research Centre, A.C.

9, Boulevard de France, bât A

1420 Braine-L'Alleud, Belgium

<http://www.iiuedu.eu/iiupress/sdsj.html>

School of Doctoral Studies (European Union) Journal

Academic Senate Delegate: *Dr. José Carlos Arias (Vice-Chancellor)*

Editor: *Dr. Anne D. Surrey*

Associate Editors: *Dr. Mathew Block*

Professor Robert Munier

Professor Alexandra Moffett

Professor Anuar Shah

Reviewers' Coordinators: *Professor Ivanna Petrovska*

Professor Melissa Brant

Professor József Ruppel

Dr. Paul Stenzel

Edition Design: *Pablo Gámez-Olivo*

Eric Veloz-Montaña

Publishers: *IIU Press and Research Centre, A.C.*

Information for Contributors

Electronic submission of manuscripts is strongly encouraged, provided that the text, tables, and figures are included in a single Microsoft Word file (preferably in Times New Roman, 12 size font).

Submit manuscript as e-mail attachment to the SDSJ Editorial Office at: edit.sdsj@iiuedu.eu. A manuscript number will be mailed to the corresponding author within the following 7 days.

The **cover letter** should include the corresponding author's full address and telephone/fax numbers and should be in an e-mail message sent to the Editor, with the file, whose name should begin with the first author's surname, as an attachment. The authors may also suggest two to four reviewers for the manuscript (SDSJ may designate other reviewers).

SDSJ will only accept manuscripts submitted as e-mail attachments.

Article Types

Three types of manuscripts may be submitted:

Regular Articles: These should describe new and carefully confirmed findings, and research methods should be given in sufficient detail for others to verify the work. The length of a full paper should be the minimum required to describe and interpret the work clearly.

Short Communications: A Short Communication is suitable for recording the results of complete small investigations or giving details of new models, innovative methods or techniques. The style of main sections need not conform to that of full-length papers. Short communications are 2 to 4 printed pages (about 6 to 12 manuscript pages) in length.

Reviews: Submissions of reviews and perspectives covering topics of current interest are welcome and encouraged. Reviews should be concise and no longer than 4-6 printed pages (about 12 to 18 manuscript pages). Reviews manuscripts are also peer-reviewed.

Review Process

All manuscripts are reviewed by an editor and members of the Editorial Board or qualified outside reviewers. Decisions will be made as rapidly as possible, and the journal strives to return reviewers' comments to authors within 3 weeks. The editorial board will re-review manuscripts that

are accepted pending revision. It is the goal of the SDSJ to publish manuscripts within the following SDSJ edition after submission.

Regular Articles

All portions of the manuscript must be typed double-spaced and all pages numbered starting from the title page.

The **Title** should be a brief phrase describing the contents of the paper. The Title Page should include the authors' full names and affiliations, the name of the corresponding author along with phone, fax and e-mail information. Present addresses of authors should appear as a footnote.

The **Abstract** should be informative and completely self-explanatory, briefly present the topic, state the scope of the work, indicate significant data, and point out major findings and conclusions. The Abstract should be 100 to 200 words in length. Complete sentences, active verbs, and the third person should be used, and the abstract should be written in the past tense. Standard nomenclature should be used and abbreviations should be avoided. No literature should be cited.

Following the abstract, about 3 to 10 **key words** that will provide indexing references to should be listed.

A list of non-standard **Abbreviations** should be added. In general, non-standard abbreviations should be used only when the full term is very long and used often. Each abbreviation should be spelled out and introduced in parentheses the first time it is used in the text.

The **Introduction** should provide a clear statement of the problem, the relevant literature on the subject, and the proposed approach or solution.

It should be understandable to colleagues from a broad range of disciplines.

Materials and methods should be complete enough to allow possible replication of the research. However, only truly new research methods should be described in detail; previously published methods should be cited, and important modifications of published methods should be mentioned briefly. Capitalize trade names and include the manufacturer's name and address. Subheadings should be used. Methods in general use need not be described in detail.

Results should be presented with clarity and precision. The results should be written in the past tense when describing author's findings. Previously published findings

should be written in the present tense. Results should be explained, but largely without referring to the literature. Discussion, speculation and detailed interpretation of data should not be included in the Results but should be put into the Discussion section.

The **Discussion** should interpret the findings in view of the results obtained in this and in past studies on the topic. State the conclusions in a few sentences at the end of the paper. The Results and Discussion sections can include subheadings, and when appropriate, both sections can be combined.

The **Acknowledgments** of people, grants, funds, etc should be brief.

Tables should be kept to a minimum and be designed to be as simple as possible. Tables are to be typed double-spaced throughout, including headings and footnotes. Each table should be on a separate page, numbered consecutively in Arabic numerals and supplied with a heading and a legend. Tables should be self-explanatory without reference to the text. The details of the research methods should preferably be described in the legend instead of in the text. The same data should not be presented in both table and graph form or repeated in the text.

Figure legends should be typed in numerical order on a separate sheet. Graphics should be prepared using applications capable of generating high resolution GIF, TIFF, JPEG or PowerPoint before pasting in the Microsoft Word manuscript file. Tables should be prepared in Microsoft Word.

Use Arabic numerals to designate figures and upper case letters for their parts (Figure 1). Begin each legend with a title and include sufficient description so that the figure is understandable without reading the text of the manuscript. Information given in legends should not be repeated in the text.

References: In the text, a reference identified by means of an author's name should be followed by the date of the reference in parentheses. When there are more than two authors, only the first author's name should be mentioned, followed by 'et al'. In the event that an author cited has had two or more works published during the same year, the reference, both in the text and in the reference list, should be identified by a lower case letter like 'a' and 'b' after the date to distinguish the works.

Examples:

Smith (2000), Wang et al. (2003), (Kelebeni, 1983), (Usman and Smith, 1992), (Chege, 1998; Chukwura, 1987a,b; Tijani, 1993, 1995), (Kumasi et al., 2001)

References should be listed at the end of the paper in alphabetical order. Articles in preparation or articles submitted for publication, unpublished observations, personal communications, etc. should not be included in the reference list but should only be mentioned in the article text (e.g., A. Kingori, University of Nairobi, Kenya, personal communication). Journal names are abbreviated according to Chemical Abstracts. Authors are fully responsible for the accuracy of the references.

Examples:

Papadogonas TA (2007). The financial performance of large and small firms: evidence from Greece. *Int. J. Financ. Serv. Manage.* 2(1/2): 14 – 20.

Mihiotis AN, Konidaris NF (2007). Internal auditing: an essential tool for adding value and improving the operations of financial institutions and organizations. *Int. J. Financ. Serv. Manage.* 2(1/2): 75 – 81.

Gurau C (2006). Multi-channel banking in Romania: a comparative study of the strategic approach adopted by domestic and foreign banks *Afr. J. Financ. Serv. Manage.* 1(4): 381 – 399.

Yoon CY, Leem CS (2004). Development of an evaluation system of personal e-business competency and maturity levels *Int. J. Electron. Bus.* 2(4): 404 – 437.

Short Communications

Short Communications are limited to a maximum of two figures and one table. They should present a complete study that is more limited in scope than is found in full-length papers. The items of manuscript preparation listed above apply to Short Communications with the following differences: (1) Abstracts are limited to 100 words; (2) instead of a separate Materials and Methods section, research methods may be incorporated into Figure Legends and Table footnotes; (3) Results and Discussion should be combined into a single section.

Proofs and Reprints

Electronic proofs will be sent (e-mail attachment) to the corresponding author as a PDF file. Page proofs are

considered to be the final version of the manuscript. With the exception of typographical or minor clerical errors, no changes will be made in the manuscript at the proof stage. Because SDSJ will be published online without access restrictions, authors will have electronic access to the full text (PDF) of the article.

Authors can download the PDF file from which they can print unlimited copies of their articles.

Copyright

Submission of a manuscript implies: that the work described has not been published before (except in the form of an abstract or as part of a published lecture, or thesis) that it is not under consideration for publication elsewhere; that if and when the manuscript is accepted for publication, the authors agree to automatic transfer of the copyright to the publisher.

Costs for Authors

Revision, edition and publishing costs will be totally paid by Secured Assets Yield Corporation Limited and authors' sole contribution will be providing BIS with their invaluable work.

Publication Decisions

Decisions by the editor of all submitted manuscripts reflect the recommendations of members of the Editorial Board and other qualified reviewers using a "blind" review process. Reviewers' comments are made available to authors. Manuscripts that are inappropriate or insufficiently developed may be returned to the authors without formal review for submission to a more suitable journal or for resubmission to SDSJ following further development.

Manuscripts submitted will be judged primarily on their substantive content, though writing style, structure and length will also be considered. Poor presentation is sufficient reason for the rejection of a manuscript. Manuscripts should also be written as concisely and simply as possible, without sacrificing clarity or meaningfulness of exposition. Manuscripts will be evaluated by the editor when first received, on their contribution-to-length-ratio, meaning that manuscripts which strong contributions will be assigned more pages than those making narrower contributions. Papers intended to make very extensive contributions (over 35 double-space pages, using one inch margins and Times New Roman twelve-pitch font) will, at discretion of the editor, be allotted additional space. Authors are expected to get and use feedback from colleagues prior to submitting a manuscript for formal review.

Section BME

Business Management and Economics

Section coordinated by the Department of Business Management and Economics in the School of Doctoral Studies of the European Union

Head of the Department:

Dr. Jünger Albinger

Deputy Head of Business Management:

Dr. Mathew Block

Deputy Head of Economics:

Dr. Paul Stenzel

Reviewers Body Coordinators:

Chair of Accounting and Finance Studies:

Professor Ira Joubert

Chair of Human Resources Studies:

Professor Beverly Lantin

Chair of Marketing Studies:

Professor Frans Cooper

Chair of Operations and Production Studies:

Professor Guy Leroy

Chair of Information and Knowledge Management Studies:

Professor Mitsuaki Uno

Chair of Leadership and Corporate Policy Studies:

Professor Tui Unterthiner

Chair of Change, Conflict and Crisis Studies:

Professor Takako Iwago

Chair of Mathematical and Quantitative Methods:

Professor Maria Nicklen

Chair of Microeconomics:

Professor Carlo Grunewald

Chair of Macroeconomics and Monetary Economics:

Professor Stefan L. Feinberg

Chair of International Economics:

Professor Mitchell Alvarez

Chair of Schools of Economic Thought and Methodology:

Professor Vincent Haidinger

Business Management and Economics Section Content

Article 1

Page 7

Empirical Evaluation of the Evidence of the Beneficial Influence of the Strategic Planning Process on the Overall Performance of Emerging Countries Companies

Author: Walter Gerard Amedzro St-Hilaire (PhD)

Country: Canada

Article 2

Page 17

Institutional Infrastructure and the Adoption of International Financial Reporting Standards (Ifrs) in Nigeria

Author: Iyoha, F.O. and Jafaru Jimoh

Country: Nigeria

Article 3

Page 24

Comparing Bankruptcy Prediction Models in Iran

Author: Ali Ebrahimi Kordlar and Nader Nikbakht

Country: Iran

Article 4

Page 32

The Relationship Between Service Quality and Satisfaction on Customer Loyalty in Malaysian Mobile Communication Industry*Author: Ahmed Audu Maiyaki (PhD)**Country: Nigeria*

Article 5

Page 39

Electronic Commerce, Automation and Online Banking in Nigeria: Challenges and Benefits*Author: Ojeka, Stephen A. (B.Sc, M.Sc Accounting) and Ikpefan, O. Ailemen (Ph.D, ACA, ACIB, FNIM)**Country: Nigeria*

Article 6

Page 51

A Survey on the Assessment of Iranian External Auditors' Recognition of Materiality Levels*Author: Azar Rahdarian(M.A.) and Hamed Hamedian(M.A.)**Country: Iran*

Article 7

Page 65

Small Scale Supplier Satisfaction: An Explorative Finding from Indian Manufacturing Industry*Author: Manoj Kumar Mohanty**Country: India*

Empirical Evaluation of the Evidence of the Beneficial Influence of the Strategic Planning Process on the Overall Performance of Emerging Countries Companies

Walter Gerard Amedzro St-Hilaire, Ph.D

Researcher on Strategy and Governance of Public Organizations in HEC Montréal

Researcher for the Research Center on the Governance of Natural Resources

Email: walter-gerard.amedzro@hec.ca

Abstract

This article raises the question of empirical assessment of the strategic planning process. Can you prove empirically, or highlight beyond a reasonable doubt, the positive influence of the process on emerging countries businesses' performance? If yes, what explanations to give to any differences of opinion between theoretical models and practical models? The first part assesses major studies based on measures of corporate performance in the emerging countries as an indicator of the effectiveness of the process. Then, the second part provides an overview of studies measuring the actual characteristics of the process. It is, thus, shown that, contrary to what is claimed by theories, the strategic planning process doesn't always lead emerging countries companies to success. Finally, the question of empirical assessment of the strategic planning process remains important because, even though it may equip these companies some advantages, there is still to prove, beyond a reasonable doubt, the cause-effect relationship between the strategic planning process and their success. However, this article highlights boundaries for the theoretical findings.

The search of an evidence of the positive influence of the strategic planning process on the overall performance of emerging countries companies led to numerous studies, since the establishment of strategic planning assumptions* by Igor Ansoff in his famous book Corporate strategy: year analytic approach to business policy for growth and expansion (1965) [1]. In this work, Ansoff said that the strategy is the key to the whole conduct of these companies' quest for openness, and in resources acquisition, distribution and processing. Combined with a rational analysis, planning is an established procedure to produce consistent results, a procedure consisting of decomposing a process in stages. It therefore makes an elaborated, formal and fragmented mechanism, a detailed plan closely linking the strategy formation process to a series of more operational steps. Thus, strategic planning would enable companies in emerging countries to improve their overall performance to cope with rapid environmental changes, implement organizational change, solve problems related to human resources management, including resistance to change. Theoretically, Strategic planning has been recognized

by many stakeholders in industrial relations, working in an organizational setting, as a process for determining the main directions of these organizations, giving these businesses the means to evolve in their environment, facing the change. But what about in practice? Do the existing strategic planning processes actually lead emerging societies to rigorous planning allowing emerging countries companies to efficiently achieve the goals they have set? Moreover, does the recent adoption of strategic planning in business modify significantly the performance of those companies? Does strategic planning increase the amount of management effectiveness of those companies? In other words, does the application of strategic planning lead, in practice, businesses to success?

These questions raise the problem of empirical assessment of the strategic planning process. Can you prove empirically, or highlight beyond a reasonable doubt, the influence of process on emerging countries businesses performance? If yes, what explanations to give to any differences of opinion between theoretical models and practical models? A review of the major available empirical studies, conducted to determine the impact and usefulness of the process allows us to make an empirical assessment of the strategic planning process. Two approaches are generally found in various studies: a method based on an

*For a complete description of the strategic planning process, consult Ansoff (1965) Corporate strategy: an analytic approach to business policy for growth and expansion

estimate of the impact of process on business performance, and a method evaluating the actual characteristics of the process. The present empirical evaluation will be done in two stages. Because evaluation of the characteristics of the process is a relatively more recent method and is based on a review of the alternative method, the first part of our evaluation will present and assess the main studies based on measures of corporate performance as index of the effectiveness of the process. Several studies have indeed been made to measure the impact of strategic planning on various aspects of business performance (financial, strategic change efficiency, overall performance ...). Then the second part of the evaluation will provide an overview of studies measuring the actual characteristics of the process. At first glance, the conclusions emerging from empirical studies on the usefulness of strategic planning are quite mixed.

Theoretical Findings

Following the success since its establishment in the 1960s, strategic planning suffered a decline in popularity and influence in the late 1970s. This was due in large measure to the failure of strategic planning tools to deliver what was expected of them. However, during the 1990s, it regained some reputation and influence that it had previously lost. Strategic planning is a powerful concept used in the business community in emerging countries, as part of a decision to deal with strategic issues. These are defined as developments, events, directions or trends that may affect the strategy of emerging countries organizations. In these organizations, strategic issues can translate individual concerns in organizational actions. They can thus be considered as having political as well as informational consequences. These effects may in turn affect these businesses decision making and strategic change in these organizations. In response to changing organizational conditions internal to these firms, and external environmental variables, the result of strategic planning is, in theory, a viable alternative, allowing these organizations to ensure the plan continually realigns the firm's goals and strategies with the changing conditions. Thus, the identification and exploitation of future opportunities, the use of strategic planning would enable these companies' major decisions to be made more efficiently and be more related to objectives. It would also allow for better allocation of time and resources to identified opportunities, and avoid wasting time and resources related to correcting erroneous or ad hoc decisions. While promoting the creation of a framework for internal communication between staff, it

allows identification of priorities within the time allotted by the plan. Ultimately, strategic planning provides a competitive advantage over these firms' competitors.

Strategic planning is therefore a tool to manage environmental turbulence. The literature describes strategic planning as an effective tool in relation to its contribution to emerging countries companies performance, or final results that the plan was originally supposed to achieve. These results are generally established at the outset by the system of strategic planning as a range of social objectives. In their research, many studies are based on the difference between formal strategic planning (or explicit) and implicit strategic planning. The formal strategic planning is an explicit and continuous organizational process, with several components (including the establishment of goals, generation and evaluation strategies). Some authors as Ansoff and Steiner suggested that a system of effective strategic planning must be linked to long term strategic objectives with those of medium-term and operational planning. Thus, planners collect data, make predictions, model and construct alternative future scenarios. These are the activities enabling these organizations to outperform those who are not engaged in a strategic planning process. But this view is not universally shared. Henry Mintzberg argues that strategic planning can be done objectively only in the short term, due to budgetary constraints, the inability to predict the future, and lack of objectivity schedules, generally biased towards the vision and desire their designers want them to take, and hierarchy[9]. With environmental constraints, emerging countries organizations need change constantly. Their definition cannot be made after an assessment of strengths and weaknesses of these organizations, but rather gradually. Strategic planning based on the needs of these organizations must respond to these needs and must take account of their gradual evolution.

For Mintzberg, the true role of strategic planning is to develop and articulate the consequences of a pre-existing strategy: planning does not create the strategy. He further argues that strategic planning is an adaptive process that evolves gradually emerging from different directions following these organizations' commitment to the environment. According to him, this emerging process is what will take these companies to success. These organizations therefore have no need for explicit planning to be a barrier to their expansion. Managers should focus more efforts on the development and optimization of capacity and efficiency of these organizations, rather than on planning, which is likely to be limiting. Ansoff, by cons, rather think that a place must be given to emerging

strategies because they are part of the explicit planning, as unexpected expenses are included in a fixed budget. This brief overview shows a theoretical divergence of theoretical perspectives on the usefulness and relevance of strategic planning. So it is important to check whether the empirical results are more in favour of one or the other models.

Empirical Evaluation

Models Based on a Measure of Performance as an Indicator of the Effectiveness of Strategic Planning Process

The main question behind these studies is: 'Does a better strategic planning result in higher levels of company performance?' To answer this question, many researchers have attempted to quantify the impact of strategic planning on the success of emerging countries' companies. In doing so, they chose different economic or financial variables (according to studies) and quantitatively measured them. The first empirical test of the relationship between strategic planning and corporate performance has been driven by Thune and House, in 1970, who surveyed 36 companies in six industry groups. Since then tests succeed, confirming or refuting their conclusions. While some studies report a positive relationship, many do not find any quantifiable benefit, and others detect even small adverse effects and costs resulting from the strategic planning.

Studies Showing a Positive Relationship Between Strategic Planning and Corporate Performance

The main studies proving the existence of a positive relationship between strategic planning and corporate performance are those conducted by Ansoff et al. (1970), Gershefski (1970), Thune and House (1970), Herold (1972), Karger and Malik (1975), Rhyne (1986). In addition, other studies conducted in the same vein, were reported by Gordon Greenley in his article "Strategic Planning and Company Performance: An Appraisal of the empirical evidence" (1994) [8] and divided into three groups. Although, in the first group, 9 studies lead to the conclusion that there is no association between strategic planning and corporate performance, in the second group, 12 studies support the evidence of such an association and, in the third, 9 conclude that companies making strategic planning outperform those that do not (Greenley 1994).

There is a difference of methodology, sampling techniques, as well as variables of interest among different studies. As an illustration, the study of Ansoff et al. uses 13 different variables to measure the performance of 93 manufacturing companies: sales, earnings, the ratio of earnings / share, total assets, report earnings / equity, ratio of dividend / shares, stocks value, the ratio of debt / equity, the capital base, the report gains / total assets, the ratio P / E, the efficiency dividends / earnings and price / equity. To ensure good reproducibility of measurements, each variable is measured 21 times. The values of those variables for emerging countries companies with an intense strategic planning are compared with those obtained for emerging countries companies with little or no strategic planning. Except for two variables (the rate of growth in equity and asset growth), companies engaged in strategic planning outperform those that don't practice it (Ansoff et al. 1970) [3]. The confidence level of the statistical tests is $\alpha < 0.1$ or $\alpha < 0.005$. The Gershefski study compares, meanwhile, sales growth in a sample of 383 companies over a period of five years before companies adopt strategic planning, and over a period of five years after the introduction of strategic planning. The results of this comparison lead the author to the same conclusion as Ansoff et al., and indicate that strategic planning is effective. With a somewhat similar methodology, Thune and House also come the same conclusion and find that emerging countries companies making explicit strategic planning surpass their own performance after the introduction of a system of explicit planning.

Study Highlighting the Lack of Quantifiable Relationships Between Strategic Planning and Corporate Performance

Many studies such as those conducted by Grinyer and Norburn in 1975 and Kulda in 1980, report the lack of quantifiable benefits from the adoption of strategic planning. Indeed, the study by Fulmer and Rue in 1973 on 386 companies and conducting a comparative analysis of four variables of financial performance (sales growth, earnings ratio, earnings growth and total capital) by distinguishing between emerging countries companies strategic planning compelled the authors to conclude that their findings call into question the most basic assumptions on which strategic planning was established. They don't deduce that strategic planning doesn't affect the final results of these companies, but explain that their study shows no clear relationship

between strategic planning and the variables measured. In their study published in 1980, Leontiades and Tezel analyze 61 companies over a period of 6 years. The approach used was to contact the Chief Executive and Head of Planning Department of each company in order to demonstrate the importance of strategic planning on various numeric and semantic domains, in order to provide quantifiable variables that would compare the performance of emerging countries companies. Based on five variables used to evaluate the performance (rate of return on equity, the assets, changes in prices and earnings, earnings per share growth unit and per sales growth unit), the study tests several hypotheses indicating a relationship between these companies performance and strategic planning. Finding none of these hypotheses statistically significant at $\alpha = 10\%$ confidence, the authors concluded that there was no evidence of such a relationship.

Studies Exposing a Negative Effect and a Cost Associated With Strategic Planning

Although they are relatively few in number, some studies suggest a negative relationship between strategic planning and corporate performance in the emerging countries. Indeed, Whitehead and Gup found some negative effects. Some of these companies using strategic planning perform less on some measures than their competitors that don't use it. In 1983, a survey by mail of 316 companies using financial planning and 133 financial companies not using it enabled them to reach these conclusions. To ensure a representative sample, each state is represented and does not contribute to over 10% of total responses. The distinction between planners and non-planners was originally based on the respective rates of sales, earnings and rates of returns on equity. To classify firms, the authors saw as advanced planners, emerging countries organizations that have explicit and written objectives. The most advanced planners were those who had specialized departments for planning and used econometric models and regression analysis to establish projects or analysis of alternative actions. Finally, the frequency of revision of strategic plans was also taken into account for purposes of this classification.

In this classification, the overall trend is clear that the use of formal planning is related to the size of these organizations: 95% of institutions with assets of \$ 1 trillion or more used a formal planning, while only 48% of institutions with assets of \$ 50 million or less used a formal planning. For their analysis, Whitehead and Gup retained three variables to measure: the rate of return on

equity and rate of return on assets to measure profit, and the absolute growth of customer deposits. Using regression analysis to isolate the impact on performance, and analysis of variance to determine whether the observed differences were statistically significant, Whitehead and Gup studied the data from their sample. They found that emerging countries institutions using strategic planning showed lower rates of return of capital and assets than those institutions that don't. For the third variable, the planners didn't obtain a growth significantly higher than non-planners. The results were obtained with a confidence level $\alpha = 0.1$ (90%). To confirm their findings, the authors redefined the distinction between planners and non-planners, regard to market expansion, product development and services, social development and in relation to social, economic and political. The results obtained by evaluating the data according to this new classification were consistent with the initial results. A series of other tests, based on a redefinition of variables and criteria distinguishing institutions whether or not using strategic planning, once again confirmed these results. The authors came to the conclusion that their results indicated a negative relationship between strategic planning and corporate performance in the emerging countries banking sector. Not rejecting the strategic planning, they wondered about the quality of planning and the existence of any competitive advantage it gave. They concluded that planning is negatively related to performance of emerging countries companies, unless it becomes profitable in long term (longer than the duration of their study). In addition, they speculated that the absence or reduction of pressures on these organizations from their environment push them to engage in strategic planning.

Methodological Criticisms

Despite the variety and number of studies to evaluate the effectiveness of the strategic planning process, it was noted a lack of methodological rigor, more or less obvious in the different studies. The critical analysis conducted by Greenley (1986) allows highlighting each one's methodological weaknesses. Thus, there is a bias in the methodological rigor of the authors, a lack of statistical tests (to check if the difference is statistically significant) or at least their omission in the publication of results. In addition, variations between variables from one experience to another, between duration of experiences, between periods (given that each period is marked by a particular situation and history) and between size and origin of the samples do not allow comparison even if

they allow a degree of complementarity of the results. The research parameters are fundamentally different from one investigator to another. In addition, some studies are marked by the absence of proof of reproducibility of measurements. Moreover, many investigations have relied on questionnaires (Ansoff et al. 1970, for example) mailed to companies. In this regard, Grinyer and Norburn rightly observe that because the planning process is complex, and spontaneous reactions to questions are important for proper assessment, mailed questionnaires were particularly inappropriate for an adequate response on the subject.

Lesson 1

The first conclusion emerging from this analysis of studies based on a measure of performance as an indicator of the relevance of the strategic planning process is that evidence of a relationship and the nature of the relationship between strategic planning and emerging countries companies performance is still unproven. We can't comment objectively on the effectiveness or ineffectiveness of the strategic planning process as a management tool. The methodological variability characterizing the different studies limits their analysis and comparison. However, by combining different studies according to their results, one finds that those that detected a positive relationship, and cause-effect relationship between strategic planning and corporate performance, are generally older than those who found no link and those who perceived harm in using planning. This time separation is sufficient to suggest that the relationship between corporate performance and strategic planning would evolve over time? Despite the strategic planning history seeming to go the same way (the popularity decline, then an interest resurgence), the lack of methodological rigor once again doesn't confirm this hypothesis. It is rather evidence of the bias in most of the authors' studies, and the use of arbitrary attributes or variables. It also shows the inadequacy of the variables used for an objective evaluation process, and the presence of other factors than those measured. There are, indeed, a whole range of other variables that may affect these companies' performance or achievements, so that the changes detected in the performance of these companies may not have been affected, or only partially affected or affected only by strategic planning. Higher levels in achieving results are not necessarily related to the use of strategic planning. It is also possible that improved performance gives these companies the means, resources for use, or the ability to implement strategic planning in

its midst. In addition, emerging countries companies can adopt strategic planning in order to protect performance previously achieved without planning. In which case there would be a relationship between strategic planning and performance, but not a causal one. Despite the conclusions he reached with his staff, Igor Ansoff admits that a subjective evaluation of results by the management doesn't differ greatly between planners and non-planners, while an objective financial measure shows a substantial difference. [3] It is therefore difficult to define specifically the consequences of the use of strategic planning. From this point of view the analysis tends to support the model of Henry Mintzberg.

However there is always a category of potential benefits resulting from the use of strategic planning. Greenley (1986) [7] recognizes in its article profits increased by the use of strategic planning, which are inherent in the consequences of its use. He refers to them in the concept of 'intrinsic values of planning'. So there are economic advantages to the use of strategic planning in emerging countries. There remains a strong a priori that strategic planning has a major effect on emerging countries companies' performance. But does planning affect these companies' performance or does performance provide them to resources for managerial attention in strategic planning? Anyway, all the problems of measuring broad performance of these companies suggest that these results overstate the true relationship between planning and performance. Models based on measures of performance are not very suitable for defining the nature of such a relationship.

Models Based on a Measurement of Process Characteristics as an Index of Efficiency of the Process

Criticisms Raised by These Models

In contrast to studies on the effectiveness of the strategic planning process related almost exclusively to the financial performance as a gauge of the value of the planning system, these models start from the assumption that the benefits of strategic planning are related to nature of the process, and may or may not be a sufficient condition for improved performance. Strategic planning can thus be effective as a process, despite the performance achieved. Hence the importance of developing more models not based on performance of economic dominance. It is important to take into account the characteristics of the process, and the dimensions of organizational context (including

the strength and resources) in which the planning takes place. The study of V. Ramanujam and Venkatraman N (1987) establishes that the dimensions of organizational context have a dominant influence on the effectiveness of the strategic planning process. Also models based on the evaluation of characteristics of the process, consider the performance of emerging countries businesses does not look sufficiently valid to base the effectiveness of the planning process. So research conducted by Greenley in 1983 and Dyson and Foster in 1982, among others, have examined the effectiveness of the process, regard to the nature of the process itself.

Definition and Characteristics of the Strategic Planning Process

The strategic planning process is defined as all human interactions, formal and informal, taking place during the generation or the development of a strategic plan. This process fulfills both a symbolic and instrumental role. Symbolically, the strategic planning process is used for building a consensus in emerging countries organizations, providing simplified models for communication and understanding. At the instrumental level, the strategic planning process serves as a program performance, accounting for uncertainties and reducing time and cost of searching for information faced by emerging countries managers in their decision-making.

This process is characterized by clarity of planning, that is to say, a division of labour among different levels of management in the initiation, formulation, revision and implementation of plans. It is also characterized by an explicit planning. Indeed, an explicit process is a more rational system for the construction of strategic plans. The third characteristic is the diversity of the process of planning. Where diversity characterizing planning is high, there are several kinds of individuals. With this kind of strategic planning process, the multiple viewpoints and conflicting are taken into account in the identification of strategic issues and developing solutions, so that the resistance is less important. Finally, the strategic planning system must be characterized by an especially intense planning. This concept refers to the level of personal resources that the participants must devote to the process of strategic planning. It indicates the involvement of everyone and one's interest in the process.

An additional feature recognized in the strategic planning process by Falshaw and Glaister (1999) is the extent to which strategies within these organizations are the result

of a deliberate or emerging process. It is on these various characteristics that this class of study models based its research. Dutton and Duncan (1987) hypothesized that the model of strategic planning process affects systematically the occurrence and success of efforts to change policy through its effects on the content and strategic issues form.

Empirical Studies

Studies based on a measurement of process characteristics as an index of efficiency of the process, usually analyze a set of dimensions of the planning system and discuss possible relations with a set of dimensions reflecting the effectiveness of the strategic planning process. The results of these studies are mixed and can be grouped into two categories: those recognizing the effectiveness of those processes and those identifying malfunctions relating to the implementation of strategic planning.

Studies Demonstrating the Effectiveness of the Strategic Planning Process

The study by Ramanujam V. and Venkatraman N (1987) illustrates the research conducted under this model. Collecting data through questionnaires from 600 companies selected from manufactures and service companies, and collecting responses from 207 of them, the authors analyzed the characteristics of these businesses and planning systems in relation to three main dimensions reflecting the effectiveness of the system (the system capacity, the objective achievements, the relative competitive performance). Following a statistical analysis, the authors came to the conclusion that there is clearly a strong relationship between the multivariate system size and dimensions reflecting the efficiency of the system. However they were unable to determine the relative importance of the contribution of the dimensions of planning systems in emerging countries to the observed relationship, that is to say the link between cause and effect.

In addition, Greenley showed in 1986 that it may affect non-financial strategic planning, that could provide a substantial benefit to emerging countries organizations. Such benefits include the process advantages, such as the ability to identify and exploit future market opportunities, personal benefits, such as encouraging a positive attitude to change, and perspective that keeps strategic planning companies in the emerging countries synchronized with their external environment so that they can cope with changes. Planning can thus be an effective management

process, despite the performance achieved. Strategies formulation comes in practice more from deliberate process, than the emerging and adaptive process supported by Henry Mintzberg. Similarly, the responses lead to concluding dysfunctions caused by strategic planning as little or not present. Few studies in this category, however, develop an impact of strategic planning with the long or short term.

Studies Identifying Dysfunctions Associated With Strategic Planning

In their 1983 study, Bresser and Bishop [13] argue that explicit strategic planning can be dysfunctional if it introduces rigidity and encourages excessive bureaucracy. In these cases the planning results in rigidity and inflexibility of responses to the changing environment in emerging countries. Strategic planning tends to increase the need for coordination and control of the process of forming strategies, usually fluid, flexible and informal. The process tends to halt the creative thinking and promote the maintenance of old patterns or models that have proved successful. In other words, in order to maintain some control, strategic planning tends to be an exaggerated extrapolation of past and present, in the future, rather than seeking to reinvent the future. Policy makers usually assume that the future is a linear progression from the past. They set the strategies taking into account a future more or less corresponding to what one knows, or some development allowed. Strategic planning creates the illusion of certainty in a world of uncertainty, risk and constant change, without taking into account the contingencies of the environment. In relatively safe environments, free from control systems and counter democratic power allowing play of market forces, or in case of monopolies or duopolies (like the field of manufacture of civil aircraft dominated globally by Airbus and Boeing companies), this illusion doesn't pose any problems. We note thus a lack of application of processes in emerging countries companies. In addition to the challenges posed by the involvement of human resources, communication and dissemination of a common culture of these companies and the adaptation of organizational structure, the lack of flexibility in the planning and the limited vision of the future that it implies prevent the efficient implementation of the strategies it has itself helped develop. These, when implemented, are sometimes inadequate and lead to unexpected results. Strategic planning is currently unable to take into account the range of possible futures, and therefore doesn't allow establishing a certain plan for the long term.

Lesson 2

Studies based on an assessment of the strategic planning process, while taking more account of the nature of strategic planning, and the nature of the consequences of planning within emerging countries organizations, also come to mixed results. They confirm the existence of a link between strategic planning and business success, but does not specify whether or not a relationship of cause and effect. Strategic planning, when adequately used, is associated with non-economic benefits that can confer or enhance the competitive advantage of a firm, but it sometimes leads to dysfunctions (including inflexibility and rigidity) that can limit these organizations in their expansion and development. It is thus clear from these studies that strategic planning is a complex tool that cuts both ways, whose effective use is not granted.

The main criticism that can be made to models measuring the characteristics of the system to evaluate its effectiveness rests once again in the methodology. Mainly based on mailed questionnaires in view of the nature of research, these studies don't record spontaneous answers of respondents, and obtain relatively low response rate (over 600 companies, 207 responded to the study of V. Ramanujam and Venkatraman N, is therefore 34.5% response, and 113 out of 500 companies responded to the study of Glaister and Falshaw, and thus an overall response rate of 23%). The result is a poor sample representation, since, as the different authors note, the majority of responses came from large corporations, or performers of strategic planning. The samples are therefore biased in favour of this category of respondents.

Explanation of Differences Between Theoretical Models and Practical Models

How to explain the differences between theory and practice? In other words, how is that, contrary to what is claimed in the literature, strategic planning doesn't always lead emerging countries companies to success, and is sometimes associated with malfunctions? Although some of these companies are favoured by the use of strategic planning, others do not recognize the benefits. Two categories of problems related to the concept may explain these differences.

Problems Related to the Nature and Definition of the Concept

The concept of strategic planning, as theoretically developed, takes little account of the cognitive limitations of human rationality, which may in turn limit the process practical effectiveness. What makes it a flawed concept that can be inefficient on certain occasions? A major problem is the lack of a consistent and meaningful definition of what constitutes a strategic planning, meaning elements of strategic planning. This problem is reflected in the various studies by the complexity and heterogeneity of preferred definitions to distinguish emerging countries organizations conducting strategic planning and those that don't. We notice in fact that requirements vary widely from study to study to classify firms by their practice of strategic planning, as well as the variables selected to measure the intensity of strategic planning within these organizations. All of these companies are involved in planning, but they differ greatly in the extent to which they are implementing the plans, grow gradually as the environment changes, and use planning tools. Also, the definition of planning varies from one company to another. It is thus unclear whether the definition adopted by a particular company at a given moment in a given context, agrees or disagrees with the theoretical definition of the elements of strategic planning. The different degrees of planning partly explain the variability of results obtained by empirical studies.

Practical Issues Related to the Development and use of the Concept

Effective planning depends on the involvement and participation of all actors involved in the life of these organizations, including officers, employees, shareholders, customers and potential strategic partners to identify priorities for these organizations, their strengths and weaknesses, and to avoid prejudicing a sector for the benefit of another, and avoid conflicts. But strategic planning is still too often the result of a small group of people. The needs (in terms of financial resources, technical, appropriate architecture, procurement, human resources, information, organizational management...) and business priorities are defined only by this group. Also planning doesn't include a number of factors (the real needs of consumers, potential, capacity for innovation and creation of employed staff, workers' interests ...), lacks objectivity and remains focused on expectations of some individuals in these organizations. It is therefore incomplete and often inadequate, with a lack of sufficient guidance on the relative priority of the basic activities, especially in emerging countries organizations divided into multiple sectors (e.g. regional governance).

Moreover, being a long and meticulous process, strategic planning requires a high expenditure of energy and time for its implementation. Moreover, despite efforts, the process doesn't always lead to expected results. In practice it is difficult to mobilize resources (especially human) necessary for its implementation.

The structure, as well as technology, doesn't always fit a new strategy. The general structure of emerging countries companies is not suitable for the use of strategic planning. Still structured according to the needs of proven traditional activities (and taking place in relatively stable environments), these companies experience structural resistance to any planning. In addition, sufficient information and collected as part of the traditional business, with competitive behaviour, is inadequate for the development of strategic alternatives needed to plan strategically. These companies don't have the complete information necessary for effective strategic planning. Information available to these organizations affect the way they are implemented: the more the information is incomplete (which come in most cases), uncertain and unreliable, the less these companies will be tempted to incur the risks in these implementations. Moreover, in these organizations divided into sectors or compartment, the structure can create barriers between different sectors if it is inadequate. And, restructuring these organizations, or adapt their structures in response to changes in strategic planning is not always easy, especially if these companies have a considerable size.

Strategic planning introduced rational elements that break with the cultural history of companies in the emerging countries and threaten the political processes. So a conflict often arises in the workplace, between activities traditionally profitable and innovative activities. This results in resistance, sometimes followed by an abandonment of the strategic planning, which limits the effectiveness of this process. Moreover, there is a failure of leaders to formulate and implement strategies. To be effective, must be involved strategic planning and leadership skills for planning and managing the overall process of strategic change, as stated Igor Ansoff (1965). However, numerous studies attempting to establish the relationship between strategic planning and firm performance don't illuminate the efforts of planning skills and strategic change management, considering strategic planning in isolation. So we can't decide on the actual intensity of strategic planning in emerging countries businesses. If in practice, this planning is carried out in isolation, these enterprises have an incomplete tool that explains the discrepancy between the theoretical and practical models.

Another explanation for the discrepancy lies in the frequency of compilation and revision of plans. Strategic planning requires constant revision of plans and re-issues of new strategies, since these are useless whenever the historical dynamics of these organizations leads where it wants to go, or the targets are proving inadequate. These revisions are necessary to ensure the flexibility of the process. Strategic planning therefore requires constant attention no matter these companies' situation (crisis or win-win situation), i.e. time and energy. However, in practice, these organizations tend to relax their attention, or devote the energy needed by planning to other activities under the conditions of the firm. In their article, Bresser and Bishop (1983) show, based on work done previously, that not very supported planning, as well as too intense planning, lead to inter-organizational contradictions and threaten the viability of emerging countries companies. Indeed, an intense planning tends to increase the new organizational products. However they often conflicts with existing ones, and result in an increase of conflicting activities within these organizations. These problems illustrate the fact that strategic planning and intensity of its use in emerging countries businesses largely depends on their size, their resources and their sphere of activity.

Conclusion

To conclude, it is important to note that the true nature of the relationship between corporate performance in emerging countries and strategic planning is still unproven. Numerous empirical studies tend to confirm one or the other theoretical models. However, the bias introduced by the methodology limits the consideration of these different studies. Models based on an assessment of corporate performance as an indicator of the effectiveness of the strategic planning process seem particularly inappropriate for such an assessment taking into account non-financial consequences arising from the application of strategic planning. Although more recent models based on an assessment of the actual characteristics of the strategic planning process somewhat compensate for this deficiency, they are also constrained by their methodology preventing them from getting a sample and responses sufficiently representative of reality. However, incorporating the dimensions of the planning system and the internal organizational environment, these models provide a better representation of reality. Although many studies in this context tend to support the theoretical model of Igor Ansoff, some turn away again. Many authors of empirical studies have highlighted the fragility of the

strategic planning process. Indeed, on one hand it is quite difficult to demonstrate clearly the benefits, and secondly, these studies have shown that despite the considerable effort made by several emerging countries companies to prepare and develop coherent strategies, few of them are actually implemented and lead to expected changes. The use of strategic planning is not acquired, and the process does not always lead to expected results. The opinions regarding the contribution of the strategic planning process to the success of these companies remain fairly divided. Although there is a link between planning and success, it is unclear whether these organizations planning leads to success or if it is success that gives these companies the means to implement a strategic planning, which, as Mintzberg says, would then serve to articulate the consequences of an existing strategy. However, it was demonstrated that strategic planning doesn't always lead emerging countries companies to success, and it doesn't provide a systematic comparative advantage over emerging countries companies that didn't adopt strategic planning. In other words, the adoption of strategic planning doesn't always lead to careful planning. Although some studies lead to believe it, others have demonstrated the existence of more or less perverse effect of the adoption of strategic planning. The question of empirical assessment of the strategic planning thus remains important because, if it has been demonstrated that strategic planning equips these companies some advantages, it wasn't possible to prove beyond a reasonable doubt, the cause-effect relationship between the strategic planning process and the success of emerging countries companies. However, the results obtained allow limiting the validity of theoretical models proposed in the literature to certain of these companies and certain time periods in history. For example, the adoption of strategic planning could lead large enterprises in the emerging countries in the 1990s to success. But a generalization of the principle to any kind of business at any time and in any country couldn't rightly be issued, for the moment.

References

- Ansoff, I. *Corporate strategy; an analytic approach to business policy for growth and expansion*; (New York : McGraw-Hill, [1965].)
- Ansoff, I. « Critique of Henry Mintzberg's *The Design School : Reconsidering the Basic Premises of Strategic Management* », *Strategic Management Journal*, 12, 449-431 (1991)

- Ansoff, I. et al. « Does Planning Pay ? The effect of planning on Success of Acquisitions in American Firms », *Long Range Planning*; 3 (2), 2-7 (1970)
- Bédard G. M. et Simard G. : *À l'aube de XXIe siècle : des enjeux pour les sciences de la gestion*; Ed Guérin Universitaire; 1996
- Bresser R. K. & Bishop R. C « Dysfunctional Effets of Formal Planning : Two Theoretical Explanations » *Academy of Management Review*, 8, 588 – 599 (1983)
- Boyd, K. B. « Strategic Planning and Financial Performance : A Meta –Analytic review », *Journal of Management Studies*; 28 : 4; 353-374; July 1991
- Dutton, J., E. & Duncan R., B., « The influence of the strategic planning process on strategic change » *Strategic Management Journal*, 8, 103 -116 (1987)
- Glaister, K. W. & Falshaw R. J. « Strategic Planning : Still Going Strong ? », *Long Range Planning*; 32 (1); 107-116; (1999)
- Greenley, G. E. « Does Strategic Planning Improve Company Performance ? », *Long Range Planning*; 19 (2); 101-109; (1986)
- Greenley, G. E. « Strategic Planning and Company Performance : an Appraisal of the empirical evidence » *Scandinavian Journal of Management*; 10 (4); 383-396; (1994)
- Mintzberg, H. « The Design School : Reconsidering the Basic Premises of Strategic Management », *Strategic Management Journal*, 11, 171-195 (1990)
- Mintzberg, H., « Learning 1, planning 0 : Reply to Igor Ansoff », *Strategic Management Journal*, 12, 463-466 (1991)
- Ramanujam V., & Venkatraman N., « Planning system characteristics and planning effectiveness » *Strategic Management Journal*, 8, 453-468 (1987)
- Whitehead, D. D. and Gup, B. E. « Bank and thift probability : does strategic planning really pay? » *Economic Review Federal Reserve Bank of Atlanta*, 70, 14 – 25 (October 1985)
- Tuval Y. : *Le Défi : contourner la concurrence*; Ed Village Mondial (Paris); 2000

Institutional Infrastructure and the Adoption of International Financial Reporting Standards (IFRS) in Nigeria

Iyoha, F.O.

Lecturer, Department of Accounting
College of Development Studies, Covenant University P.M.B 1023
Ota, Nigeria
[Email: iyohafrancis@yahoo.co.uk](mailto:iyohafrancis@yahoo.co.uk)

Jafaru Jimoh

Department of Accounting Auchi Polytechnic
Auchi, Nigeria

Abstract

Nigeria has officially indicated her intention to adopt International Financial Reporting Standards (IFRS) from year 2012. Though a welcome development, it raises the question of the country's availability and readiness of relevant and appropriate institutional infrastructure. Using the perception of users and preparers of accounting information, the results of the descriptive study showed that only the professional accounting bodies have the relevant infrastructure to cope with the adoption of the IFRS. The study recommends, amongst others that the activities of the major institutions connected with the implementation of the new standard be urgently reviewed and that the education, sensitization, and communication to stakeholders of issues associated with IFRS adoption vis a vis the role of the relevant institutions should commence in earnest and finally a rigorous IFRS capacity building program should be embarked upon by all regulatory bodies, firms and training institutions in order to provide the needed manpower for IFRS implementation. **Key words:** International Financial Reporting Standard, Institutional infrastructure, Capacity building, Adoption, Nigeria Accounting Standards Board, Relevance, Reliability, Comparability.

There is a growing demand for credible financial information to meet the needs of stakeholders who have operational interest in financial reporting. Information emanating from financial reporting is regarded as credible and useful when it faithfully represents the "economic substance" of an organization in terms of relevance, reliability and comparability (Spiceland, Sepe and Tomassini (2001:36). It is in recognition of the need to have quality financial reports that the adoption of International Financial Reporting Standards (IFRS) is taking firm root among countries around the world. One of such countries is Nigeria which has officially expressed her intention to adopt IFRS as from 2012. This position, however, is without prejudice to a host of banks in Nigeria which have already started to adopt IFRS. Though this might be regarded as a welcome development, the question that begs for answer is, does Nigeria have strong institutional infrastructure to

make the transition to IFRS (a more conceptual approach to financial reporting) effective and rewarding? The question is pertinent because IFRS is an innovation which requires institutional infrastructure to succeed not only at the development phase, but also during the implementation phase. This study, with particular focus on Nigeria, is important because institutional constraints vary from country to country in their impact on organizations (Commander and Svejnar, 2007) and the situation in Nigeria cannot be taken for granted.

The objective of this article is therefore to address the question of whether Nigeria has the required institutions to support the adoption of IFRS from the perspective of users and preparers of financial reports. The remaining part of this paper is organized as follows: Section 2 describes the institutional and infrastructural/ legal requirements for financial reporting in Nigeria. Section 3 is a review of recent

literature. Section 4 describes the methods used in this study and Section 5 survey results/discussion and finally Section 6 contains the conclusion and recommendations.

Institutional and Infrastructural/Legal Requirements for Financial Reporting in Nigeria

There are a number of institutions and agencies in Nigeria which provide guidelines that determine what information and in what format such information should be included in financial reports. And who is responsible for preparing and who should check the financial statements, deadline for producing financial statements and who is to receive a copy as well as who should approve financial statements. The totality of the guidelines are generally referred to as GAAP (Generally Accepted Accounting Principle). The Nigeria GAAP relates to Nigerian laws, particularly the Company and Allied Matters Act, (CAMA), 1990 and others which are implemented and enforced by the relevant institutions.

Though IFRSs are universal standards, however, within each country, local rules still govern (to some extent) the preparation and issue of financial statement. This arises due to the fact that IFRSs do not have the force of law, and are only effective if they are adopted by the national regulatory bodies and supported by the relevant institutions. Therefore, the need to understand the contextual significance of institutional infrastructure on the adoption of IFRS in Nigeria can not be over-emphasized. To investigate the perception of users and compilers on the adoption of IFRS in Nigeria, the study focused on the following institutions whose provisions have implications for accounting practice in Nigeria. The accountancy bodies in Nigeria-The Institute of Chartered Accountants of Nigeria (ICAN) and the Association of National Accountants of Nigeria (ANAN), the Central Bank of Nigeria (CBN), Nigerian Accounting Standards Boards (NASB), National Insurance Commission (NAICOM), Security and Exchange Commission (SEC) and Corporate Affairs Commission(CAC)

Accountancy Bodies(ICAN and ANAN)

The Institute of Chartered Accountants of Nigeria (ICAN) and the Association of National Accountants of Nigeria (ANAN) are the two bodies responsible for the production of professional accountants in Nigeria. They are also involved in ensuring that members maintain high professional conduct in the discharge of their professional duties through continuing professional

education programmes and ethical awareness. The two bodies are relevant to the adoption of IFRS in the context of the observation by Hassab et al (2001:27), that a well developed accounting profession and system of accounting education in a given country “lead to a tradition and/or effort of providing adequate reporting and disclosure.” Based on ICAN’s membership of IFAC and the number of years both bodies (ICAN and ANAN) have existed and the number of accountants produced thus far, they should be considered developed enough to impact the adoption of IFRS in Nigeria.

Central Bank of Nigeria (CBN)

The main statutory regulator of banks and non-banking financial institutions under the terms of the Banks and Other Financial Institutions Act, BOFIA (1991) is the Central Bank of Nigeria. The BOFIA contains provisions on financial reporting by banks in addition to CAMA requirements. The Act requires banks to submit audited financial statements to the CBN for approval before publication in the relevant media within four month of year-end. Based on these provisions and in addition to its role as a major regulator of the economy and with due compliance, the CBN is expected to positively impact the adoption of IFRS in Nigeria.

Securities and Exchange Commission (SEC)

The Securities and Exchange Commission (SEC) and the Nigerian Stock Exchange (SE) regulate market participants under the Investment and Securities Act of 2007 and the Securities and Exchange Rules and Regulations (2007). The development of stock market significantly influences the accounting environment of any country especially developing countries (Hassab et al, 2001). Therefore, stock market creates the need to improve corporate disclosure. Hence, the Investment and Securities Act requires every market participant to maintain accurate and adequate records of its affairs and transactions. Thus, audited financial statements must be filed with SEC and NSE before publication within three months after year-end. Based on the foregoing, the place of the SEC in the adoption of IFRS in Nigeria is paramount.

Nigerian Accounting Standards Board (NASB)

The Nigerian Accounting Standards Board sets local accounting standards under the Nigerian Accounting

Standards Board Act, 2003. Establishing and maintaining appropriate accounting standards are critical to the development of accounting practice because accounting standards allow for a more accurate reflection of the business environment by ensuring that relevant information grounded in reliable financial reporting are available to investors. Under the IFRS regime, the NASB is expected to give backing to the enforcement of the application of the standards. Based on its current structure, the extent to which its activities will be relevant under the IFRS regime is a question that begs for answer.

National Insurance Commission (NAICOM)

The NAICOM is responsible for the administration and enforcement of the provisions of the Insurance Act. To that extent, the National Insurance Commission regulates financial reporting practices of insurance companies under the National Insurance Act 2003. Audited financial statements are expected to be submitted to the NAICOM within six months of year-end and published in the relevant media. Besides, the auditor is legally required to certify the solvency of the insurer and approve the margin of solvency required under the Act. Based on its provisions with regard to accounting practice, to what extent will its activities be relevant under IFRS dispensation.

Review of Related Literature

The decision to adopt IFRS by some countries arises from the understanding that IFRS is a product with net work effect. Network effect is said to exist where users find a product or service more valuable as additional users use the same product or service. As more and more countries adopt IFRS, it becomes more appealing to others that are yet to consider the adoption. Notwithstanding that a number of challenges have been observed and experienced by countries in their decision to adopt IFRS, its worldwide adoption has been promoted on the premise of its perceived benefits which are considered to outweigh the costs.

Proponents of the adoption of IFRS argue that there are a number of benefits which can be gained from greater cross-country comparability of firms' financial reports. Barth, (2007), for instance, argues that by adopting a common body of international standards, countries can expect to lower the cost of information processing and auditors of financial reports can be expected to become familiar with one common set of international accounting standards than with various local accounting standards. If adopting IFRS

is expected to lower such costs, then Nigeria which depend to some degree on foreign capital would be interested. The argument here is that countries choose to adopt IFRS when they expect to increase the share of foreign capital and trade in their economy. In this sense, even countries with low levels of foreign capital and trade can choose to adopt IFRS if they are expecting growth in those factors.

On the contrary, opponents argue that one single set of accounting standards cannot reflect the differences in national business practices arising from differences in institutions and cultures (Armstrong, Barth, Jagolinzer, and Riedl, 2007; Access Bank, 2010). In countries where the quality of governance institutions is relatively high, IFRS adoption is likely to be less attractive as high quality institutions represent high opportunity and switching costs to adopting international accounting standards. However, in many developing countries, the quality of local governance institutions are low and thus are important determinants of the decision to adopt IFRS (Ball et al., 2000; Leuz et al., 2003; and Ball, 2006). Such countries are likely to suffer from corrupt, slow-moving, or ineffectual governments that are resistant to or incapable of change (La Porta et al., 1999). In these countries, the opportunity and switching costs are lower and thus, the chance to adopt an externally developed body of accounting standards presents an advantage. Thus, among countries with less developed institutions like Nigeria, the decision to adopt IFRS is likely to be driven by lower opportunity and switching costs.

Institutions

The key drivers of financial reporting outcomes have been observed to be institutional reporting incentives rather than accounting standards. This assertion is affirmed by recent studies which have established that institutions affect firms' actual reporting and disclosure practices irrespective of the accounting standards in force (Fan and Wong, 2002; leuz et al. 2003 and Haw et al. 2004). Therefore, the role of institutions in realizing the benefits claimed for IFRS adoption cannot be ignored as they (institutions) influence organizational choices by defining which actions are suitable and supportable in an economy. Thus, North (1990) defines institutions as the "humanly devised constraints that shape human interaction" that provide the "rules of the game in society" where the actions of players are governed by rules. Though institutions could be inefficient sometimes, they exist primarily to curtail the opportunistic tendencies which could arise in an exchange relationship. This scenario will have some implications for

IFRS adoption in that evidence support the fact that where firms are expected to apply the same accounting standards, differences in reporting practices have been observed across firms and countries (Burgstahler et al. 2006 and Lang et al. 2006).

The above exposition suggests that mandatory accounting rules and regulations such as the IFRS cannot be considered in isolation of other relevant institutions because its (IFRS) effectiveness will depend on the understanding of the economic and institutional factors that affect firms reporting incentives (Ball, 2001).

Research Methods

This is a descriptive study in which questionnaire was used to elicit data in order to determine the perception of compilers and users of annual reports on the state of institutional infrastructure for the adoption of IFRS in Nigeria. Finance managers represented the compilers because they are deemed to be knowledgeable about IFRS and could provide a preparer’s perspective on its adoption. Similarly, Investment analysts represent the users based on a number of reasons identified in extant literature. For instance, Investment analysts are identified in the literature as the principal users of financial reports (Paudyal and Rees, 2000; Healy and Palepu, 2001; Clement and Tse, 2003; Mangena, 2004) because their work requires that they have the accounting knowledge to enable them to analyze the reports and make decisions (Baker, 1998). Thus, it is believed that provision of information that meets the needs of the analysts is considered as also meeting most of the needs of other users (Gebhardt, Reighardt, and Wittenbrinck, 2004).

The questionnaire was administered to a total of 203 compilers who met the conditions specified below- The firm in which they are employed is listed and active on the Nigerian Stock Exchange in the last ten years i.e (between 1990 and December, 2009) and the industry that the firms belong has at-least three firms. Also, the questionnaire was administered to each of the Principal Analysts of the two hundred and twenty five (225) investment analyst firms identified from the list of Capital Market Operators compiled by the Nigerian Stock Exchange. All the investment analyst firms were included in the sample because, as observed by Mangena (2004:34), the response rate among investment analysts is usually low.

The questions required the respondents to answer on a five-point Likert scale anchored by (5) indicating very strong and (1) very weak. Similar scales have been used

by Firer and Meth (1986), Courtis (1992) and Myburgh (2001) and were found suitable. A total number of 125 questionnaires were returned by the compilers out of which 106 were useable. For the users, 98 questionnaires were returned out of which 75, were found useable.

Survey Results and Discussion

The results presented in the table below are based on the authors’ field survey in 2010. The first question probed the readiness of the professional accounting bodies to meet the challenges posed by the adoption of IFRS in five areas shown in the table. The results for this question are shown in Table 1.

Table 1 Professional Accounting Bodies (ICAN and ANAN)

		Compilers n=106		Users n=75	
		Mean	Rank	Mean	Rank
1	Average yearly production of accountants	3.59	2	3.10	4
2	Mandatory Continuing Professional Education (MCPE)	3.62	1	3.58	1
3	Professional ethics	3.39	3	3.46	2
4	Discipline of members	3.39	3	3.26	3
5	Service to the public and government	3.37	5	2.68	5

The results presented in the above table indicate that the education and training of accountants by the professional accounting bodies will have strong implications for IFRS. This is closely followed by the yearly production of professional accountants with a mean of 3.59 from the compilers’ point of view and professional ethics from the view point of users. These issues are important for the successful adoption of IFRS. Based on the results above, it can be inferred that on the average, the accounting bodies have the capacity to cope with the demands of IFRS adoption.

The second question probed the readiness of CBN to cope under IRFS dispensation based on its antecedents in terms of the matters listed in table 2 below.

Table 2 Central Bank of Nigeria (CBN)

		Compilers n=106		Users n=75	
		Mean	Rank	Mean	Rank
1	Submission of audited financial statements by banks	3.60	1	3.70	1

		Compilers n=106		Users n=75	
		Mean	Rank	Mean	Rank
2	Special examination of banks' books and affairs	2.24	5	3.0	4
3	Dealing with contravention of legislations and regulations	2.35	4	3.05	3
4	Maintenance of required reserves	3.40	3	2.85	5
5	Appointment of auditors and remuneration	3.50	2	3.50	2

From the table above, both compilers and users are unanimous that the CBN has been strong in ensuring submission of audited financial statements and appointment of auditors. However, CBN has been weak in the special examination of banks' books and dealing with contravention of legislations and regulations. The impression created by the results indicate that the CBN is not quite set to face the challenges of IFRS and thus would need to strengthen all areas of its mandate in order to cope with the demands of IFRS..

The third question sought to ascertain the benefits of IFRS adoption to policy makers from the stand point of the preparers and users of financial reports. The results are shown in Table 3.

Table 3 Securities and Exchange Commission

		Compilers n=106		Users n=75	
		Mean	Rank	Mean	Rank
1	Supervision of securities and market operations	2.21	6	2.50	2
2	Better access to the global capital markets	2.56	4	1.32	6
3	Promotion of cross-border investment	2.58	2	2.38	4
4	Better information for control and decision making purposes	2.60	1	2.34	7
5	Discipline of erring capital market operators	2.57	3	2.80	1
6	Approval of audited financial statements	2.48	5	2.50	2
7	Response to filing of financial statements by companies	2.10	7	2.35	5

Though the results indicated in the table above suggest that the compilers and users are not unanimous on any particular issue, but it is evident that the SEC is not strong enough as an institution to support the adoption of IFRS. The results show that SEC lack capacity in all the matters contained in the table. Overall, the results indicate that the SEC has much to do in order to improve and support the adoption of IFRS.

The fourth question probed the perception of compilers and users concerning the state of NASB towards the adoption of IFRS. The results are shown below.

Table 4 Nigerian Accounting Standards Board

		Compilers n=106		Users n=75	
		Mean	Rank	Mean	Rank
1	Improved regulatory oversight and enforcement	2.50	5	2.05	7
2	A higher standard of financial disclosure	3.60	1	3.40	4
3	Better information for market participants	2.50	5	3.56	2
4	Better ability to attract and monitor listings by foreign companies	2.60	3	3.30	5
5	Availability of resources to fulfill mandate	2.45	7	3.50	3
6	Composition of membership of the Board	3.50	2	3.60	1
7	Relationship with other Accounting bodies/institutions	2.55	4	3.00	6

The results showed in the above table indicate that the NASB lacks the necessary resources to fulfill its mandate. To that extent, the Board will lack the ability to enforce the use of IFRS. Overall, the table suggests that the NASB has a lot of areas where improvements are needed for IFRS adoption to be meaningful in Nigeria. These areas include improved regulatory oversight and enforcement, relationship with other accounting bodies and institutions and better supply of information to market participants.

The fifth question sought to ascertain the perception of the preparers and users with respect to the factors that could make the adoption of IFRS successful. Six factors which could potentially lead to successful adoption of IFRS were considered and the results are shown below.

Table 5 National Insurance Commission

		Compilers n=106		Users n=75	
		Mean	Rank	Mean	Rank
1	Having technology in place to support the conversion	3.46	3	3.41	2
2	General enforcement of the Insurance Act, 2003	3.50	2	3.35	3
3	Submission of financial statements within 6months.	2.80	5	2.65	5
4	Compliance with Solvency requirements	3.56	1	3.52	1
5	Professional support with IFRS experience	2.78	6	2.64	6
6	Sound system of corporate governance	3.38	4	3.23	4

Field survey, (2010)

The results presented above indicate that the NAICOM is perceived higher in the area of ensuring compliance with solvency requirements. However, in terms of professional support with IFRS experience and submission of financial statements within stipulated period, the Commission is weak and not likely to cope under full IFRS regime. Overall, the results indicate NAICOM is not set to embrace the adoption of IFRS.

The final question probed the perception of preparers and users in terms of the issues that will pose general challenges to the adoption of IFRS. The results are as presented in table 6.

Table 6. CHALLENGES OF IFRS ADOPTION

		Compilers n=106		Users n=75	
		Mean	Rank	Mean	Rank
1	Training of relevant personnel	3.40	6	3.45	6
2	Funding of the road map and implementation	3.85	5	3.65	5
3	Compliance and enforcement of IFRS when in force	4.45	3	4.34	2
4	Complexity of conversion from Nigeria GAAP to IFRS	4.20	4	3.89	4
5	Retention of key employees with IFRS competences	4.55	2	4.20	3
6	Ethical environment prevailing in Nigeria	4.70	1	4.60	1

The results presented in Table 6 show that the preparers and users differ only in their ranking of the challenges of the adoption of IFRS in two out of six issues raised. They are unanimous that the ethical environment is the most important challenge for the successful adoption of IFRS. This should not be surprising because corporate transparency is a particularly important component of good governance as it ensures the protection of parties (both individual and institutional) who have operational interest in financial reporting in terms of accurate and reliable information which are needed in order to take well-considered economic decisions.

Interestingly, preparers and users are both unanimous that training will be the least challenge. This is curious especially that the country (Nigeria) does not have enough trained and professionally literate IFRS accountants. One would have expected training to create the greatest obstacle especially in the issue of first time adoption of IFRS. The result also show that retention of qualified employees will

be a challenge as it is ranked second and third respectively by the preparers and users. In this early stage of IFRS, poaching of staff, especially accounting personnel might be pronounced due to the dearth of such personnel.

Conclusion

This study has provided evidence of the potential preparedness of the institutions surveyed towards the adoption of IFRS in Nigeria which has officially been scheduled to commence in 2012. The results show that only the professional accounting bodies (ICAN and ANAN) are set to embrace the demands of the introduction of IFRS in Nigeria. The other institutions included in the study have a lot to put in order if their relevance in the adoption of IFRS is to have any meaningful positive impact.

Besides the ill-preparedness of most of the institutions in Nigeria to brace up for the adoption of IFRS, there are a number of general challenges to be faced in the process of adoption of the new standard. These, among others, include ethical environment and the ability of firms to protect qualified and competent employees from being poached by other companies. Against the backdrop of the objectives of this paper, the following recommendations are hereby suggested: (a) As the time table for the adoption of IFRS in Nigeria has been determined, the need to properly reevaluate the activities of the major institutions connected with the implementation of the new standard should be urgently considered. (b) Nigeria's adoption of IFRS should be supported as a matter of national urgency to enable full attainment of the country's economic potential (c) The Nigerian Accounting Standards Board (NASB) should expedite action on the approvals and processes required for formal adoption of IFRS as national accounting standard in Nigeria (d) The education, sensitization, and communication to stakeholders of issues associated with IFRS adoption vis a vis the role of the relevant institutions should commence in earnest. (e) A rigorous IFRS capacity building program should be embarked upon by all regulatory bodies, firms and training institutions in order to provide the needed manpower for IFRS implementation.

References

Access Bank Plc (2010; Tuesday) Communique. BusinessDay 9 (186), 34-37.

- Armstrong, C., M. Barth, A. Jagolinzer, and E. Riedl. (2007). Market reaction to the adoption of IFRS in Europe. Working paper, Stanford University.
- Baker, R. G. (1998). The market for information evidence from finance directors, analysts and fund managers. *Accounting and Business Research*, 29(1), 3-20.
- Ball, R. (2006). International Financial Reporting Standards (IFRS): pros and cons for investors. *Accounting & Business Research* 36: 5-27.
- Ball, R., S.P. Kothari, and A. Robin. (2000). The effect of international institutional factors on properties of accounting earnings. *Journal of Accounting and Economics* 29: 1-51.
- Barth, M.E. (2007). Research, standard setting, and global financial reporting. Hanover, MA: New Publishers.
- Burgstahler, D. and Leuz, C. (2006). The importance of reporting incentives: Earnings management in European private and public firms. *The Accounting Review*, 81, 983-1016.
- Clement, M. B., & Tse, S.Y. (2003). Do investors respond to analysts' forecast revisions as if forecast accuracy is all that matters? *The Accounting Review*, 78(1), 227-249.
- Commander, S. and Slejnar, J. (2007). Do institutions, ownership, and reporting competition explain firm performance? Evidence from 26 transition countries. University of Michigan, Working paper.
- Courtis, J. K. (1992). The reliability of perception-based annual report disclosure studies, *Accounting and Business Research*, 23(89), 31-43.
- Fan, J. and Wong, T. (2002). Corporate ownership structure and informativeness of accounting earnings in East Asia. *Journal of Accounting and Economics*, 33. 401-425.
- Hassab, H.R., R.W. Epps and A.A.Said. (2001), The impact of environmental factors on accounting development: An Egyptian longitudinal study. Paper presented at the American Accounting Association, USA.
- Haw, I.B., Hu, L. and Wu, W. (2004). Ultimate ownership, income management and legal and extral-legal institutions, *Journal of Accounting Research*, 42, 423-462.
- Healy, P.M. & K.C. Palepu (2001). Information asymmetry, corporate disclosure and the capital markets: a review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31, 405-440.
- Lang, M., Lins, K and Miller, D. (2003). ADR, analysts and accuracy: Does cross-listing in US improve a firm's information environment and increase market value? *Journal of Accounting Research*, 4, 317-345.
- La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny. 1999. The quality of government. *Journal of Law, Economics and Organization* 15: 222-279.
- Leuz, C., Nanda, D & Wysocki. P. (2003). Earnings management and institutional factors: an international comparison. *Journal of Financial Economics*, 29, 34-67.
- Mangena, M. (2004). On the perceived importance of disclosure items in Uk interim financial reports: Evidence from the investment analysts. Working paper series, Bradford University.
- Myburgh, J.E. (2001). The informativeness of voluntary disclosure in the annual reports of listed industrial companies in South Africa. *Meditari Accountancy Research*, 9, 199-216.
- North, D. (1990), *Institutions, institutional change and economic performance*. Cambridge, UK, Cambridge University Press.
- Spiceland, J.D, J.F. Sepe and L.A. Tomassini (2001). *Intermediate Accounting*. New York. McGraw-Hill.

Comparing Bankruptcy Prediction Models in Iran

Ali Ebrahimi Kordlar

Faculty of management, Tehran university, Tehran, Iran

Nader Nikbakht

Faculty of management, Tehran university, Tehran, Iran

[Email: Nader.nikbakht66@gmail.com](mailto:Nader.nikbakht66@gmail.com)

Abstract

This paper compares the predictive power of a number of previous research models on bankruptcy prediction in Tehran Stock Exchange (TSE) from 2001 to 2009. To compare the predictive power of these models, the adjusted R² (as In-sample metric) and root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percent error (MAPC) (as out-of-sample prediction metrics). Finally, based on the estimation results of previous models in Iran, we present a final version logit \uparrow type model that has higher performance than other models. The empirical results of In-sample and out-of-sample prediction power indicate that our presented model has higher adjusted R² and lower RMSE, lower MAE and lower MAPC, too. **Key words:** Bankruptcy prediction model, Logit \uparrow model, Young test, Tehran Stock Exchange. **JEL classification codes:** M41, C23

Since Beaver (1966), an expanded literature on bankruptcy prediction has emerged, and its impact has spilled into the commercial world, where it has been used in the development of several commercially employed bankruptcy prediction models. Out of this literature have come a number of competing empirical models with alternative explanatory variables and alternative statistical methodologies for model estimation.

The dependent variable in these models is commonly a dummy variable where "firm filed for bankruptcy" is set to 1 and other is set to 0. The independent variables are often accounting ratios extracted from financial statements and include measures of profitability, liquidity, and leverage. Some studies also include market-based variables such as the volatility of stock returns and past excess returns. The accounting-based models developed by Altman (1968) and Ohlson (1980) have emerged as the most popular bankruptcy prediction models and are often used by empirical accounting researchers as indicators of financial distress.

Altman (1968) employs multivariate discriminate analysis (MDA) on a list of financial ratios to identify those ratios that are statistically associated with future bankruptcy. Ohlson (1980) uses a logit \uparrow model, which uses less restrictive assumptions than those taken by the MDA approach. Zmijewski (1984) adopts a probit approach that is also based on accounting data but uses a different set of independent variables. All of these approaches predict

future bankruptcy based on accounting ratios drawn from firm's financial statements.

More recently, Shumway (2001) has proposed a discrete-time hazard \uparrow model to predict a firm's bankruptcy using both accounting and market variables. The main difference between this model and the static logit \uparrow model is that the hazard model can be estimated within the logit \uparrow framework while using the entire life span of information (all firm-years) for each firm. By contrast, the static logit \uparrow model can only incorporate one firm-year for each observation (i.e., each observation consists of a single set of variables observed at a single point in time).

Another stream of the bankruptcy prediction literature focuses on market-based information. For example, Hillegeist et al. (2004) have developed a BSM-Prob bankruptcy prediction model that is based on the Black-Scholes-Merton option-pricing model. Their results indicate that the BSM-Prob model outperforms the models of Altman (1968) and Ohlson (1980) in a series of tests.

There are also a number of papers that propose various firm-characteristics that may be useful to future bankruptcy prediction. For instance, Rose (1992) presents a model of firm diversification in which managers use diversification to reduce the risk of bankruptcy, particularly where the ratio of the manager's firm-specific human capital to his non-firm-specific human capital is high. Denis et al. (1997) measure corporate diversification by the number of business segments. Beaver et al. (2005) propose that, other things

equal, large firms have a smaller probability of bankruptcy and that a part of this explanation is related to corporate diversification. That is, corporate diversification and firm-size are two firm-characteristics that may help to predict future bankruptcy.

Hillegeist et al. (2004) compare the performance of their BSM-Prob model against the Altman and Ohlson models in a series of in-sample and out-of-sample tests, concluding that the BSM model outperforms the accounting-based models. Similarly, Chava and Jarrow (2004) examine the relative performance of Shumway's hazard model against the Altman and Zmijewski models, concluding that the hazard model outperforms static logit models.

Wu, Gaunt and Gray (2010) build a new model comprising key variables from each of the previous models (Altman 1968; Ohlson 1980; Zmijewsky 1984; Shumway 2001; and Hillegeist et.al 2004) and add a new variable that proxies for the degree of diversification within the firm. The degree of diversification is shown to be negatively associated with the risk of bankruptcy. This more general model outperforms the existing models in a variety of in-sample and out-of-sample tests.

Base on Zmijewski and Shirata models, Mehrani, Mehrani, Monsefi and Karami (2005) present a new prediction model to bankruptcy prediction in Tehran Stock Exchange (TSE). To reach this model, research hypotheses are divided to two groups. The first group is on the power of classifying corporations in to bankrupt and non-bankrupt firms. These hypotheses confirm the power of correct classifying corporations in two bankrupt and non-bankrupt groups. The second group is on the difference between importances of financial ratios as an independent variable. These hypotheses confirm the difference between importances of independent variables in predicting corporate failures.

Etemadi, Anvariostami and Dehkordi (2009) investigate the application of Genetic Programming (GP) for bankruptcy prediction modeling. They use GP to classify 144 bankrupt and non-bankrupt Iranian firms listed in Tehran stock exchange (TSE). Then a multiple discriminant analysis (MDA) was used to benchmarking GP model. Genetic model achieved 94% and 90% accuracy rates in training and holdout samples, respectively; while MDA model achieved only 77% and 73% accuracy rates in training and holdout samples, respectively.

Rahnamai Roud Poshti, Alikhani and Maranjouri (2009) compare the results of prediction power of Altman and Fulmer models and find that there is a significant difference

between the performances of two models and show that Altman model has done more conservatively than Fulmer model as well.

Makian, Almodaresi and Karimi (2010) find that an Artificial Neural Network (ANN) model performs much better than the discriminant analysis and logistic regression techniques. Moreover, the results confirm that the accuracy of ANN model is higher than the discriminant analysis and logistic regression techniques for predicting of bankruptcy. The analysis also shows that none of the firms will bankrupt in the year after the period covered in this study. Komijani and Saadatfar (2007) find that in comparison with other models, applying neural network models can improve the potentials of financial managements to stand against economic fluctuations and bankruptcy.

The rest of paper is organized as follows. Research models and data collection process are described in section 2. Section 3 presents descriptive statistics, correlation analysis and regression results. The last section of study provides conclusion remarks.

Methodology

Research Models

There are a number of key models that have been developed by various researchers and presented in the bankruptcy prediction literature over the last three decades. A number of the most prominent models in bankruptcy prediction models are as follows:

a. Altman (1968) multiple-discriminate analysis model

$$Failing_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + \beta_5 X_{5it} + \varepsilon_{it} \quad (1)$$

Where:

Failing: is 0 for failed firm-years and 1 for other firm-years.

X1: Net working capital/total assets

X2: Retained earnings/total assets

X3: Earnings before interest and taxes/total assets.

X4: Market value of equity/book value of total liabilities.

X5: Sales/total assets.

b. Ohlson (1980) logit model

(2)

$$Failing_{it} = \left\{ 1 + \exp \left[- \left(\begin{aligned} &\alpha + \beta_1 OLSIZE_{it} + \beta_2 TLTA_{it} + \beta_3 WCTA_{it} \\ &+ \beta_4 CLCA_{it} + \beta_5 OENEG_{it} + \beta_6 NITA_{it} \\ &+ \beta_7 FUTL_{it} + \beta_8 INTWO_{it} + \beta_9 CHIN_{it} + \varepsilon_{it} \end{aligned} \right) \right] \right\}^{-1}$$

Where:

OLSIZE: Log (total assets/GNP price-level index).

The index assumes a base value of 100 for 1988.

TLTA: Total liabilities divided by total assets.

WCTA: Working capital divided by total assets.

CLCA: Current liabilities divided by current assets

OENEG: 1 if total liabilities exceed total assets, 0 otherwise.

NITA: Net income divided by total assets.

FULT: Funds provided by operations (income from operation after depreciation) divided by total liabilities.

INTWO: 1 if net income was negative for the last 2 years, 0 otherwise.

CHIN: $(NI_t - NI_{t-1}) / (|NI_t| + |NI_{t-1}|)$, where NI_t is net income for the most recent period. The denominator acts as a level indicator. The variable is thus intended to measure the relative change in net income.

The variable of Failing is defined similar to previous models.

c. Zmijewski (1984) probit model

(3)

$$Failing_{it} = \Phi(\alpha + \beta_1 NITL_{it} + \beta_2 TLTA_{it} + \beta_3 CACL_{it} + \varepsilon_{it})$$

Where:

NITL: Net income divided by total liabilities

CACL: Current assets divided by current liabilities.

and *TLTA* is defined in Ohlson's model.

d. Shumway (2001) hazard model

(4)

$$Failing_{it} = \left\{ 1 + \exp \left[- \left(\begin{aligned} &\alpha + \beta_1 NITL_{it} + \beta_2 TLTA_{it} + \beta_3 RESIZE_{it} \\ &+ \beta_4 LEXRETURN_{it-1} \\ &+ \beta_5 LAGSIGMA_{it-1} + \varepsilon_{it} \end{aligned} \right) \right] \right\}^{-1}$$

Where:

RESIZE: Log (the number of outstanding shares multiplied by year-end share price then divided by total market value).

LEXRETURN: Cumulative annual return in year t-1 minus the value-weighted TSE index return in year t-1.

LAGSIGMA: Standard deviation of the residual derived from regressing monthly stock return on market return in year t-1.

and other variables are defined in previous models.

Now, Based on previous models, we present a logit type combined model as follows:

e. Combined model

(5)

$$Failing_{it} = \left\{ 1 + \exp \left[- \left(\begin{aligned} &\alpha + \beta_1 X2_{it} + \beta_2 X3_{it} \\ &+ \beta_3 OENEG_{it} + \beta_4 INTWO_{it} \\ &+ \beta_5 NITL_{it} + \beta_6 CACL_{it} + \varepsilon_{it-1} \end{aligned} \right) \right] \right\}^{-1}$$

Where, all variables are defined in previous models. We expect that the predictive power of this model is higher than that of other models.

To compare the predictive power of research models to future bankruptcy, we use the adjusted R2 and related Young (1989) test (as In-sample prediction metric) and root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percent error (MAPC) (as out-of-sample prediction metrics)

Data

We use the 2009 version of Tadbirpardaz (the Iranian database of Tehran Stock Exchange) annual data files and sample all firms in Tehran Stock Exchange between 2001 and 2009 with sufficient data available to calculate the research variables. In some cases whereby the required data is incomplete, we use the manual archive in the TSE's library. We eliminate banks and financial institutions from sample. Imposing all the data-availability requirements yields 1,532 firm-years over the period 2001–2009. This is the full sample that we use for testing research hypotheses.

Research results

Descriptive Statistics

Table 1 provides descriptive statistics of research variables in failed and continuous firm-years. There are 142 failed firm-years and 1390 continuous firm-years. In our available statistical population the mean (median) of $X1$, $X2$, $X3$, $X4$ and $X5$ in failed firms are -0.50 (-0.38), -0.63 (-0.42), -0.07 (-0.05), 0.40 (0.27) and 0.61 (0.53). The mean and median of all these variables for failed firms are lower than the mean and median of mentioned variables for continuous firms. The mean (median) of $OLSIZE$ for failed firms, -0.71 (-0.77) is lower than that of $OLSIZE$ for continuous firms, -0.07 (-0.15), too. Failed firms on average have a higher debt ratio (1.31) than continuous firms (0.65). Also, the mean of ratio of current assets on current liabilities ($CLCA$) for failed firms (1.77) is greater than this ratio for continuous firms (0.95). On average,

failed firms (-3.72) are smaller than the continuous firms (-3.10). The mean of $WCTA$, $CLCA$ and $TLMTA$ for failed firms (0.67, 1.77 and 0.77, respectively) is higher than the mean of these variables for continuous firms (0.60, 0.95, and 0.50, respectively).

The mean of $NITA$, $FULT$, $CHIN$, $NITL$ and $CACL$ (-0.14, 0.01, -0.05, -0.10 and 0.68) for failed firms is lower than the mean of mentioned variables for continuous firms (0.13, 0.22, 0.03, 0.23 and 1.19, respectively). Furthermore, the mean (median) of $LEXRETURN$ and $LAGSIGMA$ are -0.05 (-0.02) and 0.10 (0.03), respectively and are lower than the mean (median) value of these variables for continuous firms, 0.18 (0.15) and 0.12 (0.10), respectively.

Correlation Analysis

The Pearson correlation coefficients between research variables are provided in Table 2. In this table, the coefficients that are marked with asterisk (*), are not significant at the ordinary significance level.

Table 1. Descriptive statistics

	Failed firms: 142 Obs.						continuous firms: 1390 Obs.				
	Mean	Median	Maximum	Minimum	St.Dev		Mean	Median	Maximum	Minimum	St.Dev
$X1$	-0.50	-0.38	0.21	-2.63	0.52		0.04	0.04	0.49	-0.69	0.17
$X2$	-0.63	-0.42	-0.05	-2.97	0.59		0.11	0.08	0.50	-0.20	0.11
$X3$	-0.07	-0.05	0.17	-0.43	0.10		0.16	0.14	0.55	-0.19	0.10
$X4$	0.40	0.27	5.47	0.03	0.64		1.68	0.93	15.51	0.03	2.05
$X5$	0.61	0.53	1.85	0.02	0.34		0.83	0.78	2.59	0.04	0.37
$OLSIZE$	-0.71	-0.77	2.69	-3.20	1.16		-0.07	-0.15	4.41	-3.20	1.32
$TLTA$	1.31	1.19	3.34	0.57	0.51		0.65	0.66	1.20	0.19	0.15
$WCTA$	0.67	0.70	0.93	0.20	0.17		0.60	0.62	0.94	0.11	0.19
$CLCA$	1.77	1.52	5.49	0.64	0.91		0.95	0.88	4.95	0.26	0.39
$NITA$	-0.14	-0.12	0.19	-0.62	0.13		0.13	0.11	0.53	-0.19	0.10
$FULT$	0.01	0.00	0.41	-0.26	0.11		0.22	0.17	1.53	-0.36	0.26
$CHIN$	-0.05	-0.02	0.94	-1.00	0.43		0.03	0.05	1.00	-1.00	0.31
$NITL$	-0.10	-0.10	0.25	-0.26	0.09		0.23	0.16	1.52	-0.24	0.23
$CACL$	0.68	0.66	1.57	0.18	0.27		1.19	1.14	3.84	0.20	0.42
$RESIZE$	-3.72	-3.76	-2.25	-4.67	0.48		-3.10	-3.13	-1.35	-4.70	0.66
$LEXRETURN$	-0.05	-0.02	0.57	-1.00	0.22		0.18	0.15	1.46	-1.77	0.33
$LAGSIGMA$	0.10	0.03	0.68	0.00	0.13		0.12	0.10	0.74	0.00	0.09
$TLMTA$	0.77	0.79	0.97	0.15	0.15		0.50	0.52	0.97	0.06	0.21

Table 2. Pearson correlation coefficient

	X1	X2	X3	X4	X5	OLSIZE	TLTA	WCTA	CLCA	NITA	FUTL	CHIN	NITL	CACL	RESIZE	LEXRETURN	LAGSIGMA
X2	0.74																
X3	0.42	0.60															
X4	0.18	0.26	0.54														
X5	0.17	0.15	0.29	0.01*													
OLSIZE	-0.04*	0.15	0.09	-0.08	-0.15												
TLTA	-0.80	-0.86	-0.56	-0.35	-0.07	-0.07											
WCTA	0.31	-0.12	-0.11	-0.24	0.26	-0.17	0.24										
CLCA	-0.90	-0.61	-0.37	-0.18	-0.23	0.08	0.67	-0.39									
NITA	0.51	0.68	0.91	0.55	0.20	0.12	-0.68	-0.18	-0.44								
FUTL	0.19	0.36	0.61	0.57	0.12	0.02*	-0.43	-0.32	-0.20	0.56							
CHIN	0.06	0.09	0.29	0.16	0.09	0.01*	-0.07	0.00*	-0.07	0.33	0.10						
NITL	0.39	0.52	0.81	0.64	0.13	0.04*	-0.59	-0.23	-0.37	0.86	0.69	0.28					
CACL	0.76	0.44	0.35	0.29	0.16	-0.15	-0.57	0.31	-0.77	0.39	0.28	0.07	0.51				
RESIZE	0.09	0.28	0.45	0.43	-0.07	0.76	-0.26	-0.27	-0.04*	0.48	0.30	0.10	0.40	-0.01*			
LEXRETURN	0.10	0.16	0.27	0.21	0.01*	0.15	-0.09	0.04*	-0.09	0.29	0.03*	0.09	0.20	0.06	0.34		
LAGSIGMA	0.10	0.05	0.05	0.02*	0.03*	0.05	-0.06	0.03*	-0.10	0.07	-0.04*	0.06	0.01*	0.04*	0.11	0.09	
TLMTA	-0.34	-0.41	-0.67	-0.79	-0.07	0.10	0.54	0.28	0.30	-0.71	-0.56	-0.16	-0.73	-0.37	-0.51	-0.28	-0.08

This table presents Pearson correlation coefficient. The coefficients that are marked with asterisk (*), are not significant but others are significant at the 5% level or lower.

The results of Pearson correlation coefficients show that the most of coefficients between research variables are significant at the 5% level or better.

Model Estimation

The regression results of models and In-sample and out-of-sample prediction metrics to compare the models

are presented in Table 3. The estimation results of Altman (1968) model indicate that X1 (-0.12), X2 (-0.48) and X3 (-0.52) have a negative and significant relationship with dependent variable (Failing) at the 1% level and X4 (0.01) has a positive relation with dependent variable at the 5% level but X5 has no significant relationship with dependent variable. The F-statistic of this model (468.16) is significant at the 1% level.

Table 3: The results of regression models

	Altman (1968) model		Ohlson (1980) model	Zemijewsky (1984) model	Shumway (2001) model	Combined model
Panel A: Model estimation results						
Intercept	0.18**		-12.87**	-8.00**	-16.94**	-2.22**
X1	-0.12**					-0.95
X2	-0.48**					-13.40**
X3	-0.52**					
X4	0.01*					
X5	0.00					
OLSIZE			-0.14			
TLTA			14.58**	8.21**	17.75**	
WCTA			-2.92			
CLCA			-0.60			

	Altman (1968) model		Ohlson (1980) model	Zemijewsky (1984) model	Shumway (2001) model	Combined model
OENEG			-0.17			1.73**
NITA			-1.40			
FUTL			-2.60			
INTWO			3.79**			0.25
CHIN			-0.25			
NITL				-3.21*	-9.88*	-0.91
CACL				-0.03		0.01
RESIZE					0.01	
LEXRETURN					-1.25	
LAGSIGMA					-0.60	
TLMTA						
Panel B: In-sample prediction power metric						
Adj. R ²	51.25	Pse. R ²	83.21	72.71	76.07	87.83
F stat.	468.16**	LR stat.	904.61**	1272.92**	1083.25**	1384.16**
Young Z	11.48**		2.21*	4.81**	3.22**	
Panel C: Out-of-sample prediction power metrics						
RMSE	0.21		0.14	0.16	0.14	0.12
MAE	0.14		0.03	0.06	0.05	0.03
MAPE	4.96		1.64	2.69	2.05	1.41

*, ** significant at the 5% and 1%, respectively

In the Ohlson (1980) model, TLTA (14.58) and the dummy variable of INTWO (3.79) are significant at the 1% level and in Zemijewsky (1984) model TLTA (8.21) and NITL (-3.21) are significant at the 1% and 5% level, respectively. The results of Shumway (2001) model indicates that TLTA (17.75) and NITL (-9.88) are significant at the 1% and 5% level respectively. Finally, our presented model show that X2 (-13.40) and OENG (1.73) are both significant at the 1% level.

The results of likelihood ratio (LR) for Ohlson model (904.61), Zemijewsky model (1272.92), Shumway model (1083.25) and our presented combined model (1384.16) show that all models are significant at 1% level, generally.

The In-sample prediction power (adjusted R²) of our combined model (87.83%) is higher than that of Altman model (51.25%), Ohlson model (83.21%), Zemijewsky model (72.71%) and Shumway model (76.07%). the results of Young (1989) test show that the differences between the prediction power of combined model and Altman model (11.48), Ohlson model (2.21), Zemijewsky model (4.81) and Shumway model (3.22) are all significant at the 1% level.

The results show that the root mean squared error (RMSE) of our combined model (0.12) is lower than that of Altman model (0.21), Ohlson model (0.14), Zemijewsky model (0.16) and Shumway model (0.14). The results

also indicate that the mean absolute error (mean absolute percent error) of our combined model, 0.03 (1.41) is lower than (or equal to) that of Altman model, 0.14 (4.96), Ohlson model, 0.03 (1.64), Zemijewsky model, 0.06 (2.69) and Shumway model, 0.05 (2.05). Thus, the results indicate that the In-sample and out-of-sample prediction power of our presented model in predicting failed firms is higher than that of other previous models. Thus, our logit type bankruptcy prediction model significantly outperforms other models.

Conclusion

In this paper, we examine the predictive power of a number of bankruptcy prediction models. The models use a range of different independent variables (accounting information and market and firm-characteristic data) and a range of different econometric specifications (multiple-discriminate analysis, logit, probit models). We find that firms are more likely to experience bankruptcy if they have relatively lower earnings before interest and tax to total assets, a larger decline in net income, relatively low working capital to total assets, or high total liabilities to total assets.

We compare the empirical performance of a range of bankruptcy prediction models using a series of in-sample and out-of-sample performance metrics. Across all of these

metrics, both in-sample and out-of-sample, we find that, our combined model significantly outperforms models from the extant literature.

References

- Agarwal, V., and Taffler, R. (2008), "Comparing the performance of market-based and accounting-based bankruptcy prediction models", *Journal of Banking & Finance* 32, PP: 1541-1551.
- Altman, E. I. (1968), "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy", *Journal of Finance* 23, PP: 589-609.
- Beaver, W. H. (1966), "Financial ratios as predictors of failure", *Journal of Accounting Research* 4, PP: 71-111.
- Beaver, W. H., McNichols, M.F., and Jung-Wu, R. (2005), "Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy", *Review of Accounting Studies* 10, PP: 93-122.
- Begley, J., Ming, J., and Watts, S. (1996), "Bankruptcy classification errors in the 1980s: an empirical analysis of Altman's and Ohlson's models", *Review of Accounting Studies* 1, PP: 267-284.
- Burgstahler, D., James, J., and Noreen, E. (1989), "Changes in the probability of bankruptcy and equity value", *Journal of Accounting and Economics* 11, PP: 207-224.
- Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008), "In search of distress risk", *Journal of Finance* 63, PP: 2899-2939.
- Chava, S., and Jarrow, R. A. (2004), "Bankruptcy prediction with industry effects", *Review of Finance* 8, PP: 537-569.
- Denis, D. J., Diane, K., and Denis, A. S. (1997), "Agency problems, equity ownership, and corporate diversification", *Journal of Finance* 52, PP: 135-160.
- Dichev, I. D. (1998), "Is the risk of bankruptcy a systematic risk?", *Journal of Finance* 53, PP: 1131-1147.
- Etemadi, H., Anvari Rostami, A. A., and Dehkordi, H. F. (2009), "A genetic programming model for bankruptcy prediction: empirical evidence from Iran", *Expert Systems with Applications* March, 36, PP: 3199-3207.
- Francis, J. (1990), "Corporate compliance with debt covenants", *Journal of Accounting and Research* 28, PP: 326-347.
- Hillegeist, S. A., Keating, E., Donald P., Cram, J. G., and Lünstedt. (2004), "Assessing the probability of bankruptcy", *Review of Accounting Studies* 9, PP: 5-34.
- Komijani, A., and Saadatfar, J. (2007), "Application of "artificial neural network models", in predicting economic bankruptcy of the registered corporations in stock market", *Journal of Economic Essays* 3, PP: 11-44.
- Lyandres, E., and Alexei, Z. (2007), "Investment opportunities and bankruptcies prediction. Working Paper, Rice University, Houston. Ohlson, James, 1980. Financial ratios and the probabilistic prediction of bankruptcy", *Journal of Accounting Research* 18, PP: 109-131.
- Makian S.N.A.D., Almodaresi S.M.T., and Karimi Takalou S. (2010), "A comparison among artificial neural network, discriminant analysis and logistic regression techniques for bankruptcy: a case study of Kerman's firms", *The Economic Research* 10, PP: 141-161.
- Mehrani, S., Mehrani, K., Monsefi, Y., and Karami, Gh. R. (2005), "Implication Zmijewski and Shirata bankruptcy prediction models in Tehran stock exchange", *The Iranian Accounting and Auditing Review* 12, PP: 105-131.
- Rahnamai Roud Poshti, F., Alikhani, R., and Maranjouri, M. (2009), "Applicational investigation of Altman and Fulmer bankruptcy prediction models in Tehran stock exchange", *The Iranian Accounting and Auditing Review* 16, PP: 19-34.
- Rose, D. (1992), "Bankruptcy risk, firm-specific managerial human capital, and diversification", *Review of Industrial Organization* 7, PP: 65-73.
- Shumway, T. (2001), "Forecasting bankruptcy more accurately: a simple hazard model", *Journal of Business* 74, PP: 101-124.
- Stone, M. (1991), "Firm financial stress and pension plan continuation/replacement decisions", *Journal of Accounting and Public Policy* 3, PP: 175-206.
- Vassalou, M., and Xing, Y. (2004), "Default risk in equity returns", *Journal of Finance* 59, PP: 831-868.
- Vuong, Q. H. (1989), "Likelihood ratio tests for model selection and non-nested hypotheses", *Econometrica* 57, PP: 307-333.

- Wu, Y., Gaunt, C., and Gray, s. (2010), "Acomparison of alternative bankruptcy prediction models", *Journal of Contemporary Accounting and Economics* 6, PP: 34-45.
- Zmijewski, M. E. (1984), "Methodological issues related to the estimation of financial distress prediction models", *Journal of Accounting Research* 22 (Suppl.), PP: 59-82.

The Relationship Between Service Quality and Satisfaction on Customer Loyalty in Malaysian Mobile Communication Industry

Sany Sanuri Mohd. Mokhtar

Chairperson of Quality at Quality Management Centre and Senior Lecturer at the College of Business, University, Utara Malaysia (UUM). PhD in Marketing and Quality Management. Sany was formerly a deputy director of Quality Management Institute, UUM.

Ahmed Audu Maiyaki (corresponding author)

Ph.D. Candidate in Marketing at the College of Business, Universiti Utara Malaysia.

Lecturer at Bayero University Kano, Nigeria.

T 11 008, Maybank Residential Hall Universiti Utara Malaysia, 06010 Sintok

College of Business, Universiti Utara Malaysia

[Email: aamaiyaki@yahoo.com](mailto:aamaiyaki@yahoo.com)

Norzaini bt Mohd Noor

Masters degree holder from UUM. Researcher on service quality and service marketing.

Abstract

The paper explores the relationship between service quality and customer satisfaction on customer loyalty with regards to mobile phone usage among the postgraduate students of a university in Northern Malaysia. The sample consists of 341 students randomly selected from the population frame provided by the university. The respondents are made of up students from various part of the world such Asia, Middle-east and Africa. The results show that both service quality and customer satisfaction significantly affect the level of customer loyalty of mobile phone users in Malaysia. Hence, all the hypotheses except one have been supported. It was therefore, recommended that mobile service providers should pay special attention to their service quality and the factors that drive customer satisfaction. Suggestion for future research was also offered. **Key words:** Service quality, Customer satisfaction, Loyalty, Mobile communication, Malaysia.

Over the last few years, the number of mobile phones has increased at an exponential rate globally. This increase is more pronounced in the developed countries. The reasons for this are numerous, low acquisition price and availability of the equipment contributes in no small measure to its widespread usage. Furthermore, the number of mobile phones in the world has already passed the number of fixed land lines and the revenue from mobiles phones will soon exceed that of fixed land lines. In the era of improved mobile communication technology, vast amount of changes are generated in facilitating communication and the transfer of information from business to business, business to customers, employers to employees among others. Consequently, the utilization of mobile phones in

communication and information transfer leads to providing more and more added value services (Steenderen, 2002). Despite the various information services provided through mobile phone services nowadays, detail assessments need to be made in order to understand the needs and requirements of the mobile phone users.

Service quality and customer satisfaction are inarguably the two core concepts that are at the crux of the marketing theory and practice (Spreng & Mackoy, 1996). In today's world of intense competition, it is generally believed that the key to sustainable competitive advantage lies in delivering high quality service that will in turn result in satisfied customers. The prominence of these two concepts is further manifested by the root of theoretical and empirical

studies on the topic that have emanated over the past few years. Therefore, there is no doubt about the importance of service quality and customer satisfaction as the ultimate goals of service providers. To this end therefore, the study seeks to understand the nature of mobile phone use among respondents at a higher learning institution and investigate their perceptions on the mobile phone applications in the context of services quality.

Literature Review

Service Quality

Parasuraman, Berry and Zeithaml (1991) defined service quality as “the extent of discrepancy between customer’s expectations or desires and their perceptions.” After in-depth interviews of executives and a few focus groups in four service categories, Parasuraman, Zeithaml and Berry (1985) developed a service quality model, which identify five gaps between customers and marketers. The gap model of service quality was first developed by Parasuraman et al. (1985). According to Ueno (2010) the authors observe that the service quality gaps model is the conceptualisation of service quality as the gap between customer expectations on the services and perceptions of the service performance. Furthermore, based on the gap model there are four major discrepancies contributing to service quality gaps, and each of the four gaps in turn contributes to the existence of gap 5 (Ueno, 2010).

Gap 1: The first gap is between consumer expectations and management perceptions of consumer expectations. In the Parasuraman et al. (1985) study, the authors found that privacy or confidentiality during transactions emerged as a pivotal quality attribute in every banking and securities brokerage focus group; however this considerations was mentioned rarely by the executives. Parasuraman et al. (1985) concluded that the lack of understanding of this gap will have an impact on the consumer’s evaluation of the service quality.

Gap 2: The second gap is between management perceptions of customer expectation and service quality specifications. Even though the executives attempt to match or exceed customer expectations, they find it difficult to deliver what the customer expects (Parasuraman et al., 1985). The authors indicated that one of the reasons for this is the difficulty in establishing specifications to deliver a fast response consistently because of a lack of trained service personnel and the wide range of function demand.

Gap 3: The third gap is between service quality specifications and service actually delivered. High service quality cannot be guaranteed even when there are guidelines for carrying out excellent services. Parasuraman et al. (1985) indicated that the employees of a service company play an important role on the service quality and the employee’s performance cannot always be standardized.

Gap 4: The fourth gap is between service delivery and what is communicated to customers about the service. Parasuraman et al. (1985) contended that promising more than can be delivered has a detrimental effect on customer because it raises the initial expectations but lowers perception quality.

Gap 5: The fifth gap is between the customer’s perceptions of service quality and their expectations of service quality. Parasuraman et al. (1985) discovered that the key to service quality is to meet or exceed the expectations of customers.

Subsequently, based on the gap model in their exploratory study, Parasuraman et al. (1985) came up with ten dimension of service quality namely: reliability, responsiveness, competence, access, communication, courtesy, credibility, security, understanding and tangibles. However, these ten dimensions of service quality were collapsed into 5 dimensions due to overlaps. The five dimension of service quality according to Parasuraman, Zeithaml and Berry (1988) are: tangibility, reliability, assurance, responsiveness and empathy. Although, criticisms have been made against the measurement of service quality “SERVQUAL” developed by Parasuraman et al. (1988), their contribution in the area of service quality continue to be significant for over quarter a century.

Customer Loyalty

Loyalty of customers is considered to be a function of satisfaction and that loyal customers contribute to company profitability by spending more on company products and services, via repeat purchasing, and by recommending the organization to other consumers (Bowen & Chen, 2001; Fecikova, 2004). To further understand the behavior of loyal customers, recent research has attempted to integrate the concept of customer commitment (Fullerton, 2005; Zins, 2001). For the most part, these recent studies have been built upon customer commitment as a key mediator of the relationship between the customer's evaluations of a firm's performance and the customer's intentions regarding the future relationship with the firm (Fullerton, 2005).

Customer loyalty expresses an intended behavior related to the service or the company. This includes the likelihood of future renewal of service contracts, how likely it is that the customer changes patronage, how likely the customer is to provide positive word-of-mouth, or the likelihood of customers providing voice. If real alternatives exist or switching barriers are low, management discovers the organization's inability to satisfy its customers via two feedback mechanisms: exit and voice (Hirschman, 1970). Exit implies that the customers stop buying the company's services while voice is customer complaints expressing the consumers' dissatisfaction directly to the company. Customers' exit or change of patronage will have an impact on the long-term revenue of the company. Effects caused from changes in the retention rate are exponential (not linear) with regard to effects on the long-term revenue. Even a marginal reduction/increase in retention rate has significant effects on future revenue (Andreassen, 1995; Reichel & Sasser, 1990).

Three conceptual perspectives have been suggested to define customer loyalty: the behavioral perspective, the attitudinal perspective and the composite perspective (Bowen & Chen, 2001; Zins, 2001). The behavioral perspective, "purchase loyalty", strictly looks at repeat purchase behavior and is based on the customer's purchase history. The emphasis is on past -rather than on- future actions. Moreover, no other loyal behavioral actions such as price tolerance, word of mouth, or complaint behavior can be interpreted (Zins, 2001). Concentrating on the behavioral aspect of loyalty could overestimate true loyalty (Zins, 2001). The attitudinal perspective, in contrast, allows gain in supplemental understanding of loyal behavior (Zins, 2001). The customer loyalty is approached as an attitudinal construct. An Attitude denotes the degree to which a consumer's disposition towards a service is favorably inclined. This inclination is reflected by activities such as the customers recommending service providers to other consumers or their commitment to repatronize a preferred service provider (Gremler & Brown, 1996). Based on a favorable attitude towards a service provider, customers may develop "preference loyalty" (De Ruyter, Wetzels & Bloemer, 1998). Lastly, the composite perspective combines attitudinal and behavioral definitions of loyalty. The composite perspective might be considered as an alternative to affective loyalty since using both attitude and behavior in a loyalty definition arguably increases the predictive power of loyalty (Pritchard & Howard, 1997).

Under all these conditions, protecting the existing customer base and retaining existing customer loyalty

appear to be the crucial competitive advantage. Customer loyalty is a key component for a brand's long-term viability (Krishnamurthi & Raj, 1991). Oliver (1991) defines it as "a deeply held commitment to re-buy or re-patronize a preferred product/service consistency in the future, thereby causing repetitive same-brand or same brand-set purchasing, despite situational influences and marketing efforts of having the potential to cause switching behavior." Though there are many more definitions of customer loyalty, it seems clear that there are two basic varieties: stochastic and deterministic. Fournier and Yao (1997) observe that there is a need for a complete definition of brand loyalty, identify six necessary conditions: a biased (i.e. random) behavioral response (i.e. purchase), expressed over time by some decision-making unit, with respect to one or more alternative brands out of a set of such brands, as a function of psychological processes (such as decision making or evaluation).

No matter how customer loyalty is defined, in order to gain it, any operator needs to:

- Increase subscriber satisfaction by raising offered service quality (for example, Anderson and Sullivan, 1993; Brady and Robertson, 2001; Kristensen, Martensen & Gronholdt, 2000; Fornell, Johnson, Anderson, Cha & Bryant, 1996; Oliver, 1980);
- Ensure subscribers' trust in the firm (see, for example, Fournier, 1998; Gundlach, Achrol & Mentzer, 1995; Morgan and Hunt, 1994; Lau and Lee, 1999); and
- Establish a cost penalty for changing to another service provider, making that a comparatively unattractive option (Fornell, 1992) and expand its application (Eber, 1999; Jones, Beatty & Mothersbaugh, 2002; Bloemer, De Ruyter & Peeters, 1998; Lee, Lee & Feick, 2001).

Customer Satisfaction and Customer Loyalty

Several authors have found a positive correlation between customer satisfaction and loyalty (Anderson & Sullivan, 1993; Bearden & Teel, 1980; Bolton & Drew, 1991; Fornell, 1992). Customers may be loyal because of high switching barriers or lack of real alternatives. Customers may also be loyal due to their satisfaction and thus want to continue the relationship. History has proven that most barriers to exit are limited with regard to durability; companies tend to consider customer satisfaction the only viable strategy in order to keep existing customers.

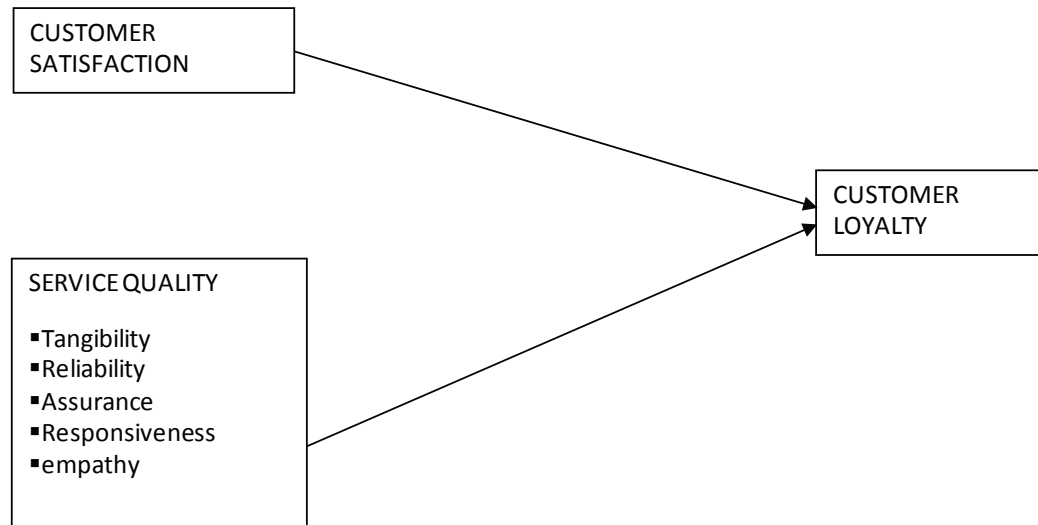
Based on Coyne (1989), there are two critical thresholds affecting the link between satisfaction and loyalty. On the high side, when satisfaction reaches a certain level, loyalty

increases dramatically; at the same time, when satisfaction declined to a certain point, loyalty dropped equally dramatically (Oliva, Oliver & MacMillan, 1992).

Managers of the company should realize that having satisfied customers is not good enough they must have

extremely satisfied customers. Moreover, a small increase in customer satisfaction can result in boosted customer loyalty dramatically.

Conceptual model



Hypothesis

From the theoretical framework discussed above, six hypotheses were developed for this research as follows:

H1: There is a significant relationship between tangibility and customer loyalty

H2: There is a significant relationship between reliability and customer loyalty

H3: There is a significant relationship between responsiveness and customer loyalty

H4: There is a significant relationship between assurance and customer loyalty

H5: There is a significant relationship between empathy and customer loyalty

H6: There is a significant relationship between customer satisfaction and customer loyalty.

Methods

The population of the study consists of all international students in Universiti Utara Malaysia amounting to 3012 see appendix. The population is made up of students from various countries, most of them come from Asia, Middle East and Africa in that order. This enables responses from people with different perceptions due to perhaps cultural variation. Based on Sekaran's (1992) recommendation, 350 respondents were randomly selected from across all the 71

courses run in the University. Similarly, the sample selection is also in line with the suggestion of Krejcie and Morgan (1970) that the good sampling number for this population size is round 341 respondents. The questionnaire used in this study consists of four sections which seek information from the respondents on: their demographic characteristics, service quality of mobile providers, customers' level of satisfaction and finally, the extent of customer loyalty. The items are measured based on 7-point Likert-type rating scale. Ultimately, the data collected was analysed using Statistical Package for Social Sciences (SPSS) version 12 for windows. More specifically, descriptive statistics, reliability test and regression analysis were run.

Results

Service Quality of Mobile Service Provider and Customer Loyalty

The positive coefficient for the service quality dimensions of tangibility, reliability, responsiveness and assurance suggests that high service quality leads to higher level of customer loyalty. However, the empathy dimensions had negative coefficient, meaning decreasing level of customer loyalty with high empathy element of service quality. In other words, higher empathy results in decreasing client loyalty. Additionally, among the variables, only tangibility,

assurance and empathy have significant values and thus, significantly contributed to the explanation of the dependent variable. To this end, hypotheses H1, H2, H3 and H4 are supported and hypothesis H5, rejected. Overall, this study concluded that mobile service provider service quality elements affects client loyalty by explaining 48 percent of the variance in the Dependent Variable (loyalty) which is quite respectable considering the nature of the research – consumer behaviour.

Table I: Regression results of Service Quality and Customer Loyalty of mobile service provider

Model	Unstandardized Coefficients		Standardized Coefficient	t	Sig.
	B	Std. Error	Beta		
1 Constant}	0.482	0.320		1.507	0.133
Tangibility	0.398	0.062	0.365	6.394	0.000
Reliability	0.103	0.089	0.074	1.154	0.249
Responsiveness	0.035	0.101	0.024	0.346	0.730
Assurance	0.519	0.089	0.376	5.866	0.000
Empathy	-0.147	0.055	-0.131	-2.673	0.008

Dependent variable: customer loyalty R^2 0.486
Adjusted R^2 0.478

Customer Satisfaction and Customer Loyalty

Table II shows the regression results indicating the relationship between client satisfaction and client loyalty. The R^2 value of 0.711 indicates that over seventy percent of client satisfaction is associated with client loyalty. Meaning that over 70% of the variance in the dependant variable (loyalty) is accounted for by customer satisfaction. In other words the positive significant coefficient suggests higher client satisfaction on mobile service provider results into higher level of client loyalty to the mobile provider. To this end therefore, hypothesis H6 is equally supported.

Table II: Regression results of customer satisfaction and customer loyalty of mobile service provider

Model	Unstandardized Coefficients		Standardized Coefficient	t	Sig.
	B	Std. Error	Beta		
1 Constant}	0.458	0.169		2.708	0.007
CS	0.916	0.032	0.843	28.910	0.000

Dependent Variable: Customer Loyalty R^2 0.711
Adjusted R^2 0.710

Discussion and Conclusion

The study found that the customer satisfaction plays an important role to enhance the level of customer loyalty. This means the higher the level of customer satisfaction the more loyal the customer of mobile phone become and the reverse is the case. This result is consistent with the findings of Cronin & Taylor (1992). The implication for the management of mobile phone service providers is that they should strive to ensure high level of customer satisfaction which will eventually leads to customer loyalty. Ensuring customer satisfaction could perhaps be achieved by identifying, focusing and improving the factors that clients consider in determining their level of satisfaction with a particular mobile service provider. The fact that this is beyond the scope of this study, we hereby recommend that future research investigate the factors determining customer satisfaction for a particular mobile service.

Furthermore, the findings also indicate that service quality has influence on customer loyalty with regards to the service delivery of mobile service providers. In fact, the findings of this studies show that service quality explained customer loyalty with values less than that of customer satisfaction i.e. 58% as against 71%. The implication of this for mobile service providers is that the elements of customer satisfaction are very important in determining the level of customer loyalty to a particular service provider. However, considering the large values of r square for both service quality and customer satisfaction there is need to pay particular attention by mobile service providers to elements of service quality and customer satisfaction. As a limitation of this study, it does not investigate the relationship between service quality and satisfaction. Therefore, the future research could explore the interrelationships between service quality and customer satisfaction.

References

- Anderson, E.W., & Sullivan, M.W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science*, Vol. 12 No.2, pp.125-43.
- Andreassen, T.W. (1995). Small, high cost countries strategy for attracting MNC's global investments. *The International Journal of Public Sector Management*, Vol. 8 No.3, pp. 24-38

- Bearden, W.O., & Teel, J.E. (1983). Selected determinants of consumer satisfaction and complaint reports. *Journal of Marketing Research*, Vol. 20, pp.21-8.
- Bloemer, J., De Ruyter, K., & Peeters, P. (1998). Investigating drivers of bank loyalty: the complex relationship between image, service quality and satisfaction. *International Journal of Bank Marketing*, Vol. 16, No.7, pp.276-86.
- Bolton, R.N., & Drew, J.H. (1991). A multistage model of customers' assessments of service quality and value. *Journal of Consumer Research*, Vol. 17, pp.375-84.
- Bowen, J.T., & Chen, S.L. (2001). The relationship between customer loyalty and customer satisfaction. *International Journal of Contemporary Hospitality Management*, Vol. 13, No.5, pp.213-7.
- Brady, M.K., & Robertson, C.J. (2001). Searching for a consensus on the antecedent role of service quality and satisfaction: exploratory cross-national study. *Journal of Business Research*, Vol. 51, pp.53-60.
- Coyne, K. (1989). Beyond service fads – meaningful strategies for the real world. *Sloan Management Review*, Vol. 30, pp.69-76.
- Cronin, J.J., & Taylor, S.A. (1992). Measuring service quality: are-examination and extension. *Journal of Marketing*, Vol. 56, pp.55-68.
- De Ruyter, K., Wetzels, M., & Bloemer, J. (1998). On the relationship between perceived service quality, service loyalty and switching costs. *International Journal of Service Industry Management*, Vol. 9, No.5, pp.436-53.
- Eber, N. (1999). Switching costs and implicit contracts. *Journal of Economics*, Vol. 69, No.2, pp.159-71.
- Fecikova, I. (2004). An index method for measurement of customer satisfaction. *The TQM Magazine*, Vol. 16, No.1, pp.57-66.
- Lee, J., Lee, J., & Feick, L. (2001). The impact of switching costs on the customer satisfaction-loyalty link: mobile phone service in France. *Journal of Services Marketing*, Vol. 15, No.1, pp.35-48.
- Fournier, S. (1998). Consumers and their brands: developing relationship theory in consumer research. *Journal of Consumer Research*, Vol. 24, No3, pp.343-73.
- Fournier, S., & Yao, J.L. (1997). Reviving brand loyalty: a conceptualization within the framework of consumer-brand relationships. *International Journal of Research in Marketing*, Vol. 14, No.5, pp.451-72.
- Fornell, C. (1992). A national customer barometer: the Swedish experience. *Journal of Marketing*, Vol. 56 No.1, pp.6-21.
- Fornell, C., Johnson, M.D., Anderson, E.W., Cha, J., & Bryant, B.E. (1996). The American customer satisfaction index: nature, purpose, and findings. *Journal of Marketing*, Vol. 60, No.10, pp.7-18.
- Fullerton, G. (2005). How commitment both enables and undermines marketing relationships. *European Journal of Marketing*, Vol. 39 No.11-12, pp.1372-88.
- Gremler, D.D., & Brown, S.W. (1996). Service loyalty: its nature, importance and implications. in Edvardsson, B., Brown, S.W., Sohnston, R., Scheuing, E. (Eds), *QUIS V: Advancing Service Quality: A Global Perspective*, ISQA, New York, NY, pp.171-81.
- Gundlach, G.T., Achrol, R.S., & Mentzer, J.T. (1995). The structure of commitment in exchange. *Journal of Marketing*, Vol. 59, No.1, pp.78-92.
- Hirschman, A.O. (1970). Exit, voice, and loyalty: responses to decline in firms. *Organizations, and States*, Harvard University Press, Cambridge, MA.
- Jones, M.A., Beatty, S.E., & Mothersbaugh, D.V. (2000). Switching barriers and repurchase intentions in services. *Journal of Retailing*, Vol. 76, No.2, pp.259-74.
- Jones, M.A., Beatty, S.E., & Mothersbaugh, D.V. (2002). Why customers stay: measuring the underlying dimensions of services switching costs and managing their differential strategic outcomes. *Journal of Business Research*, Vol. 55, pp.441-50.
- Krejcie, R. & Morgan, D. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30, 607-610
- Krishnamurthi, L., & Raj, S.P. (1991). An empirical analysis of the relationship between brand loyalty and consumer price elasticity. *Marketing Science*, Vol. 10, No.2, pp.172-83.
- Kristensen, K., Martensen, A., & Gronholdt, L. (2000). Customer satisfaction measurement at Post Denmark: results of application of the European customer satisfaction index methodology. *Total Quality Management*, Vol. 11, No.7, pp.1007-15.

- Lau, G., & Lee, S. (1999). Consumers' trust in a brand and link to brand loyalty. *Journal of Market Focused Management*, Vol. 4, pp.341-70.
- Morgan, R.M., & Hunt, S.D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, Vol. 58 No.7, pp.20-38.
- Oliva, T.A., Oliver, R.L., & MacMillan, I.C. (1992). A catastrophe model for developing service satisfaction strategies. *Journal of Marketing*, Vol. 56, pp.83-95.
- Oliver, R.L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, Vol. 17, No.4, pp.460-9.
- Oliver, R. (1997). *Satisfaction: A Behavioral Perspective on the Consumer*. New York, NY: McGraw-Hill.
- Parasuraman, A., Zeithaml, V., & Berry, L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, Vol. 49, pp.41-50.
- Parasuraman, A., Zeithaml, V.A., & Berry, L.L. (1988). SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, Vol. 64, No.1, pp.12-40.
- Pritchard, M.P., & Howard, D.R. (1997). The loyal traveler: examining a typology of service patronage. *Journal of Travelers Research*, Vol. 35, No.4, pp.2-11.
- Reichel, F., & Sasser, W.E. (1990). Zero defections: quality comes to services. *Harvard Business Review*, Vol. 68, pp.105-11.
- Sekaran, U. (1992). *Research methods for business a skill-building approach* (2nd ed.). New York : John Wiley
- Ueno, A. (2010). What are the fundamental features supporting service quality? *Journal of Services Marketing*, Vol. 24, No. 1, pp. 74–86
- Zins, A.H. (2001). Relative attitudes and commitment in customer loyalty models. *International Journal of Service Industry Management*, Vol. 12, No.3, pp.269-94.

Electronic Commerce, Automation and Online Banking in Nigeria: Challenges and Benefits

Ojeka, Stephen A. (B.Sc, M.Sc ACCOUNTING)

Accounting Department Covenant University Nigeria

[Email: stojeka@yahoo.com](mailto:stojeka@yahoo.com)

Ikpefan, O. Ailemen (Ph.D, ACA, ACIB, FNIM)

Banking and Finance Department Covenant University Nigeria

[Email: ochei_ikpefan@yahoo.co.uk](mailto:ochei_ikpefan@yahoo.co.uk)

Abstract

Electronic banking has been around for some time in the form of automatic teller machines and telephone transactions. More recently, it has been metamorphosed by the Internet; a new look and delivery channel for banking services that benefits both customers and banks. The objective of this paper is to find out the correlation between the anticipated benefits/challenges and encountered benefits/challenges. This paper therefore empirically, adopted the use of survey research to explore in quantitative terms the various challenges and benefits e-business poses to Nigeria businesses, with particular reference to Banking and Finance Industry. It was found out that there is statistically significant difference between the anticipated and encountered benefits and major challenge is the security breach faced the customers. We therefore recommend that workshops should be organized for customers periodically on how to keep their data secret especially on how to combine numbers to form password and Constant training of employees both local and international on new development in online trading should be encouraged. **Key words:** e-commerce, Electronic Banking, Automation, e-business, Banking/ Finance Industry.

E-banking includes familiar and relatively mature electronically-based products in developing markets, such as telephone banking, credit cards, ATMs, and direct deposit. It also includes electronic bill payments and products mostly in the developing stage, including stored-value cards (e.g., smart cards/smart money) and Internet-based stored value products. E-banking in developing countries is in the early stages of development. Most banking in developing countries is still done the conventional way. However, there is an increasing growth of online banking, indicating a promising future for online banking and Nigeria banks are taking good advantage of it.

Nigerian banks started very low in the quest for the adoption for electronic banking but this slow pace witnessed at the beginning of last decade is fast changing for the better in term of adoption of e-banking. Adeyemi (nd) posited that slow adoption of electronic banking practice is rapidly changing for the better. This assertion was supported by Ayo; Adebisi; Ekong & Fatudimu (2007) where they posited that with improved technological development and provision of basic infrastructure there will be improved

e-Commerce and e-Payment services with overall reduction in the amount of currency in circulation.

Awareness of electronic payments in Nigeria is increasing and it accounted for N360 billion worth of transaction in 2008 (Adeyemi, n.d). (Ayo et al., 2007 cited in Adesina & Ayo, 2010) submitted that, this revolution started in the Nigeria banking system in 2003 with the introduction of Guideline of Electronic Banking by the Central Banking of Nigeria. This was accompanied by bank reformation exercise in June 2004. The reformation exercise left Nigeria with 24 strong and reliable banks against 89 banks previously in existence. The author further maintained that, the surviving banks of the recapitalization exercise have enormously engaged the use of ICT as a platform for effective and efficient delivery of banking services. This has made Nigerian banking sector more competitive because customers are now yearning for more online services that will cater for all their needs right from the rooms with their desktops, laptops, and palmtops and even from their handsets or desks in their various offices without necessarily step into the banking hall. Global demand as a result of inflow of

cash into the economy has also been linked to this dramatic change in ICT embracement. Muniruddeen, (2007) cited in Adesina & Ayo, (2010) corroborated this submissions that the banks' huge investment in telecommunication networks and various e-Banking services delivery could be seen as an effort towards measuring up with global standard. This is among other reasons such as increased customer demand, increased competition among banks themselves; derive minimized cost, new entrants, and better service delivery. Moreso, Schaechter (2002) argued that electronic banking has made it easier for customers to compare banks' services and products, increase competition among banks, and allows banks to penetrate new markets and thus expand their geographical reach. Some even see electronic banking as an opportunity for countries with underdeveloped financial systems to leapfrog developmental stages. Customers in such countries can access services more easily from banks abroad and through wireless communication systems, which are developing more rapidly than traditional "wired" communication networks. Prior to the introduction of electronic banking in Nigeria in the 1990's masterminded by the new generation banks such as Intercontinental Bank, Zenith Bank, Guarantee Trust Bank etc., financial services delivery was very poor. Customers had to spend hour in long queues in the banking hall to carry out transactions either to withdraw or deposit cash into their account. This was the era of manual processing of transactions. The old generation banks such as United Bank of Africa, First Bank of Nigeria and Union Bank of Nigeria saw themselves as lords in the financial service industry. They dictated the pace in the banking industry and being market leader with many products and services, customers had no choice than to patronize them. With the emergence of internet and electronic banking, customers' expectations in financial services delivery are yet to be met, hence the reason for this paper. This paper therefore is poised to look at the various benefits and challenges that were anticipated before the advent of internet banking (that is, in those days when there were long queues and spending of hours in the banking hall to carry out transactions either to withdraw or to deposit cash into your account) vis-a-vis the encountered benefits and challenges after the advent of online banking. This was as a result of the existing gap between actual and expected financial services delivery to customers.

The objective is to find out the correlation between the two pairs of the variables measured i.e. the anticipated challenges and benefits with encountered challenges and benefits. But even with the development of e-commerce,

online banking in developing countries has yet to receive any significant attention among researchers, and so echoes the general lack of information systems research in sub-Saharan Africa (Mbarika, Okoli, Byrd & Datta, 2004). Hence the reason for this research works. This paper is divided into five sections. Section one above is the introduction, section two captured the literature review, section three looked at the methodology, while four and five end the paper with discussion of findings, conclusion and recommendation.

Literature Review

The concept of e-banking includes all types of banking activities performed through electronic networks. It is the most recent delivery channel of banking services which is used for both business-to-business (B2B) and business-to-customer (B2C) transactions (Mohammad, 2009). The definition of e-banking varies amongst researchers partially because electronic banking refers to several types of services through which a bank customer can request information and carry out most retail banking services via computer, television or mobile phone (Daniel 1999; Molls 1998; Sathye, 1999). Burr (1996) describes e-banking as an electronic connection between the bank and customer in order to prepare, manage and control financial transactions while Leow, Hock Bee (1999) state that the terms Personal Computer (PC) banking, online banking, Internet banking, telephone banking or mobile banking refers to a number of ways in which customer can access their banks without having to be physically present at the bank branch. Therefore, e-banking covers all these ways of banking business electronically (Mohammad, 2009). The discovery of internet and what we called electronic commerce have opened various opportunities for online trading all over the world. It has brought the market close to the customers and potential customers at a relatively low cost. Infact, online purchase reduces cost compare to physical visiting of shop for purchase. This has opened the market of the developed countries to the entire world. Also, financial institutions being the financial of the economy have been the champion of this crusade where their customers i.e. borrowers of fund cut across various countries. Adesina & Ayo, 2010 disclosed that the advent of Internet, electronic commerce, communication technology and users' response to this technology has opened opportunity for many businesses including the financial institution. Before this period, there were just a few dial-up e-mail providers in

Nigeria before 1998; a couple of Internet Service Providers (ISPs) that operated on slow links. For years, Nigerian Telecommunications (Nitel), the parastatal monopoly, dominated on the whole Nigeria telecommunications market. E-commerce in Nigeria is just at the beginning stage. From the convenient places like their homes or offices, these days most banks offer internet banking services which enable the customers to conduct banking transactions online. Today the majority of the Nigerian banks offer online, real-time banking services.

The area of e-commerce that has developed in Nigeria mostly is e-banking. A few banks started the ATM Consortium in 2003 to set up ATMs across the country. Nigeria is far behind other countries in providing technology at an affordable cost to its population. Adesina & Ayo (2010) also maintained that the adoption of electronic banking service delivery is fast gaining ground in Nigeria. Different e-Banking channels such as electronic cards, internet banking and mobile banking services have been introduced. Pikkarainen et al. (2004) highlighted two major reasons underlying online banking development and penetration. First, banks get significant cost savings in their operation through e-Banking services and secondly, that banks have reduced their branch networks and downsized the number of service staff. It was also indicated that electronic banking services delivery are the cheapest, the most profitable and wealthiest delivery channel for banking products. Internet banking services are crucial for long-term survival of banks in the world of electronic commerce (Burnham 1996). The market for internet banking is forecast to grow sharply in the next few years, affecting the competitive advantage enjoyed by traditional branch banks (Duclaux 1996; Liao, Shao & Chen, 1999). It was also argued that internet banking would help banks present a potentially low cost alternative to brick and mortar branch banking (Margaret & Thompson 2000). Nigeria Direct, (2006); Onwuka, (2006) cited in Francis & Babatunde (2009) posited that with a population of over 150 million that is growing at 3 percent annually, Nigeria has witnessed an increased demand for improved service delivery and convenience by consumers. Banks can provide improved service delivery and convenience by enhancing their value networks through online banking (Sannes, 2001; Crane, & Bodie, 1996). Francis & Babatunde, (2009) therefore submitted that some banks in Nigeria, amongst other facets of the economy, have taken advantage of the country's ICT infrastructure to improve services to customers.

Benefits of E-Commerce

The uptake of e-commerce is influenced by its potential to create business value and by awareness of its participants of the potential benefits (Salnoske, 1997). A major reason for most companies, irrespective of size, to participate in business is to extract some benefit from it. E-commerce is no different Joze, Julie & Angela (2002). The benefits of e-commerce identified from the current literature as adapted from Joze, Julie & Angela (2002) are classified in two main categories - tangible and intangible. The authors work carried out in Australian is similar to what this paper is considering in Nigeria. Hence the adaptation of the benefits and challenges as presented in the tables below. Table 1 presents the key benefits as described in the literature.

Table 1

Benefits	Research/literature
Tangible benefits	
Business efficiency	(Fraser et al. 2000; Lee 2001; Riggins, 1999) Amrit (2007)
Increased automation of processes	(Fraser et al. 2000; Dan et al, 2001)
Transformation of traditional market chain	(Fraser et al. 2000), Amrit (2007)
Retained and expanded customer base	(Fraser et al. 2000; Rahul, Biju and Abraham 2001; Turban, et al, 2000)
Reduced operation costs	(Kent and Lee, 1999; Grover and Ramanlal, 2000; Kare-Silver, 1998; Fergusson, 1999)
Acquisition of a niche market	(Riggins, 1999; Rahul et al. 2001)
Intangible Benefits	
Enhancing well-being and education of customers	(Whinston et al. 1997; Lee 2001), Amrit (2007)
Consumer loyalty	(Lee 2001; Hoffman et al. 1999; Coulson, 1999)
Competitive advantage	(Kalakota et al. 1999; Hoffman et al. 1999; Straub, 2000; Kare-Silver, 1998), Amrit (2007)
Convenient shopping	(Hannon, 1998; Winner, 1997)

Adapted from Joze, Julie & Angela (2002)

Challenges of E-Commerce

To extract benefits from e-commerce, it is important for businesses to overcome the e-commerce inhibitors and challenges Joze, Julie & Angela (2002). Findings have also shown that insecurity; inadequate operational facilities

like telecommunication and electric supply are among hindrances to online banking in Nigeria (Ezeoha 2005; Chiemeka et al., 2006). E-commerce challenges identified from the literature are classified as - technological, managerial, and business related and are summarised and presented below.

Table 2

Challenges	Research/literature
Technological challenges	
Security	(Koved et al. 2001; Czerniawska & Potter, 1998; Alexander, 1998, Chaechter, 2002)
Web site issues	(Watson et al. 1999; Zhang & von Dran 2000; Lee 2001)
Technology issues including costs, software, infrastructure	(Hoffman et al. 1999; Abeyesekera et al. 1999; Rahul et al. 2001, Chaechter 2002)
Managerial challenges	
People and organisational issues	(Hoffman et al. 1999; Feeny 2000)
Obtaining senior management backing	(Feeny 2000)
Business challenges	
Customer service	(Whinston et al. 1997; Alter, 1999; Lee 2001)
Customers' old habits	(Hoffman et al. 1999; Schwartz, 1999)
Legal issues	(Hoffman et al. 1999; Lawrence et al. 1998; de Souza & von Wiese 2000), Chaechte, (2002)

Adapted from Joze, Julie & Angela (2002)

Prior Studies on Benefits And Challenges of Online Banking

A previous study in Joze, Julie & Angela (2002) showed that that the major benefits of e-commerce adoption not anticipated by the sector are business, efficiency, improved image, competitive advantage, increased automation of processes and increased business turnover. Also, the key challenges identified for the sector are the costs of the technology, the lack of knowledge of e-commerce, managing the change, budgeting and issues associated with linking back end systems. They did not consider secure transactions as a major challenge for the sector; in contrast they were considered one of the success factors.

Pohjola (2002) also showed that the contribution of the use of information communication technology to growth of output in the Finnish market sector has increased from 0.3

percentage points in early 1990s to 0.7 points in late 1990s. Similarly, research conducted in Estonia (Aarma and Vensel, 2001), bank customers use bank office services on average 1.235 times per month, and wait in queue in bank office on average for 0.134 hours. Simple calculation shows, that making payments via E-banking facilities (for instance using Internet bank) rather than in the bank offices create overall economy savings in the amount of 0.93% of GDP (Average distance to nearest bank office is 4.14 km (Aarma and Vensel, 2001), which takes approximately 0.21 hours to travel. (BankAway, 2001; Gur_u, 2002) also considered the benefit from the customer point of view that there is a reduction in costs of accessing and using the banking services, increased comfort and timesaving - transactions can be made 24 hours a day without requiring the physical interaction with the bank, quick and continuous access to information and corporations will have easier access to information as, they can check on multiple accounts at the click of a button, better cash management

According to a survey by Booz, Allen & Hamilton (1996), an estimated cost providing the routine business of a full service branch in USA is \$1.07 per transaction, as compared to 54 cents for telephone banking, 27 cents for ATM (Automatic Teller Machine) banking and 1.5 cents for internet banking (Nathan 1999; Pyun et al., 2002). In Nordea Bank, Finland, one online transaction costs the bank an average of just 11 cents, compared to \$1 for a transaction in the branch (Echikson, 2001). Average payment in internet bank or via direct debit cost 4 times less, than payment in branch. On actual cost side (or cost side from the bank point of view), average direct debit payment cost 16 times less and payment in internet bank 7 times less, than payment in branch. Amrit (2007) however submitted that risk management, infrastructure development and policy formulation are the three major challenges of E-banking in Nepal. Technological problems like connect break in service while withdrawing cash from ATM and poor mobile service. He also considered that an adequate level of infrastructure and human capacity building are required before banks adopt the full-fledged E-banking. But Mohammad (2009) summarized the major risk of e-banking as operational risks (e.g. security risks, system design, implementation and maintenance risks); customer misuse of products and services risks; legal risks (e.g. without proper legal support, money laundering may be influenced); strategic risks; reputation risks (e.g. in case

the bank fails to provide secure and trouble free e-banking services, this will cause reputation risk); credit risks; market risks; and liquidity risks

Method

The research work was carried out in the Banking industry in Nigeria based on a sample of three deposit money banks in Nigeria. The research work was designed in such a way that data was generated from the questionnaires personally administered to the staffs of the respective banks with a high response rate of 90%. The sample size of this research work constitutes three Nigerian banks. They include: Oceanic Bank, UBA Bank and Intercontinental Bank. The three sample banks are made up of two old generation banks and one new generation bank. A total number of 120 questionnaires was administered, 40 for each of sample company. These banks were selected using judgmental sampling technique in order to have a representation of the population. Krejcie & Morgan (1970) in Amadii (2005) agrees with the sample as they proposed the population proportion of 0.05 as adequate to provide the maximum sample size required for generalization. The banks were selected because of their size and wide range of products which are all over the country. To the best of the researcher's judgment, the banks make a good representation of the banking industry in Nigeria. The expert opinion was sought for in order to validate the content and the structure of the questionnaire during the pilot study.

For testing the hypothesis, a statistical parametric test called Pair Sample t-test was employed to test the significance difference between the pre and post of online banking in Nigeria through the use of SPSS statistical package. Our intention is to establish if there is any significant difference between the anticipated benefits and challenges and encountered benefits and challenges of online banking in Nigeria.

Statement of Hypothesis

H_0 : There is no significant difference between the anticipated and encountered benefits of online banking in the Nigerian Banking system

H_0 : There is no significant difference between the anticipated and encountered challenges of online banking in the Nigerian Banking system

Decision Rule

Reject H_0 if P value is $< .05$ and accept H_0 if P value is $> .05$

Empirical Results and Implication of Findings

Expected/Encountered Benefit of Online Banking in Nigeria

Table 1. Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	INTURNEXP	3.67	90	.474	.050
	INCTURNENC	3.56	90	.500	.053
Pair 2	INCBUSACTEXP	3.50	90	.503	.053
	INCBUSACTENC	3.39	90	.594	.063
Pair 3	COMPADVEXP	2.94	90	.916	.097
	COMPADENC	3.22	90	.715	.075
Pair 4	OPCOSTREDEXP	3.06	90	.853	.090
	OPCOSTREDENC	2.94	90	.784	.083
Pair 5	EXPCUSTBASE	3.39	90	.682	.072
	CUSTBASEEXENC	3.56	90	.500	.053
Pair 6	LOYALTYEXP	3.28	90	.561	.059
	LOYALTYENC	3.06	90	.407	.043
Pair 7	REDWATIMEXP	3.39	90	.831	.088
	REDWATIMENC	2.94	90	.625	.066
Pair 8	BETTUNDEXP	3.11	90	.661	.070
	BETTUNDENC	3.11	90	.461	.049
Pair 9	INCNTWOKEXP	2.89	90	.880	.093
	INCNTWOKENC	3.11	90	.941	.099
Pair 10	POSIMAGEEXP	3.39	90	.594	.063
	POSIMAGEENC	3.28	90	.561	.059

Source: Computer Printout of Researchers' Survey

Table 2. Paired Samples Test

		Paired Differences					T	Df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	INTURNEXP – INCTURNENC	.111	.570	.060	-.008	.230	1.850	89	.068
Pair 2	INCBUSACTEXP – INCBUSACTENC	.111	.661	.070	-.027	.250	1.595	89	.114
Pair 3	COMPADVEXP – COMPADENC	-.278	.735	.077	-.432	-.124	-3.586	89	.001
Pair 4	OPCOSTREDEXP – OPCOSTREDENC	.111	1.249	.132	-.151	.373	.844	89	.401
Pair 5	EXPCUSTBASE – CUSTBASEEXENC	-.167	.691	.073	-.311	-.022	-2.288	89	.025
Pair 6	LOYALTYEXP – LOYALTYENC	.222	.632	.067	.090	.355	3.335	89	.001
Pair 7	REDWATIMEXP – REDWATIMENC	.444	1.072	.113	.220	.669	3.934	89	.000
Pair 8	BETTUNDEXP – BETTUNDENC	.000	.887	.093	-.186	.186	.000	89	1.000
Pair 9	INCNTWOKEXP – INCNTWOKENC	-.222	1.322	.139	-.499	.055	-1.595	89	.114
Pair 10	POSIMAGEEXP – POSIMAGEENC	.111	.741	.078	-.044	.266	1.422	89	.158

Source: Computer Printout of Researchers'

Expected/Encountered Challenges of Online Banking

Table 3. Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	HCOSTECHEXP	3.33	90	.750	.079
	HCOSTECHENC	3.61	90	.490	.052
Pair 2	LECOMMKWNEXP	2.83	90	.503	.053
	LECOMMKWNENC	3.06	90	.709	.075
Pair 3	ITSKCOSTEXP	2.83	90	.963	.101
	ITSKCOSTENC	2.83	90	.838	.088
Pair 4	ECOMMINFRCHEXP	2.83	90	.503	.053
	ECOMMINFRCHENC	2.83	90	.604	.064
Pair 5	RESTOMANTELEXP	2.83	90	.963	.101
	RESTOMANTELENC	3.28	90	.561	.059
Pair 6	GETRELVENEXP	3.11	90	.570	.060
	GETRELVENENC	3.06	90	.709	.075

		Mean	N	Std. Deviation	Std. Error Mean
Pair 7	SECBREACHEXP	2.94	90	.527	.056
	SECBREACHENC	3.22	90	.632	.067
Pair 8	RESTTONLBANEXP	2.56	90	.602	.063
	RESTTONLBANENC	2.61	90	.831	.088
Pair 9	NONACEORURPEXP	2.94	90	.709	.075
	NONACEORURPENC	2.94	90	.625	.066
Pair 10	NONCAPOFDATEXP	2.44	90	.689	.073
	NONCAPOFDATENC	2.78	90	.858	.090
Pair 11	MAKBUZKNTCUSTEXP	2.50	90	.963	.101
	MAKBUZKNTCUSTENC	3.22	90	.715	.075
Pair 12	UNRELOFINTPROEXP	3.11	90	.661	.070
	UNRELOFINTPROENC	2.89	90	.570	.060
Pair 13	LAKOFPROBUDGTEXP	3.11	90	.570	.060
	LAKOFPROBUDGTENC	3.17	90	.691	.073
Pair 14	POWAFALILEXP	3.06	90	.625	.066
	POWAFALILENC	2.94	90	1.085	.114

Table 4. Paired Samples Test

		Paired Differences					T	Df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	HCOSTECHEXP – HCOSTECHENC	-.278	.874	.092	-.461	-.095	-3.014	89	.003
Pair 2	LECOMMKWNEXP – LECOMMKWNEC	-.222	.790	.083	-.388	-.057	-2.668	89	.009
Pair 3	ITSKCOSTEXP – ITSKCOSTENC	.000	1.341	.141	-.281	.281	.000	89	1.000
Pair 4	ECOMMINFRCHEXP – ECOMMINFRCHENC	.000	.581	.061	-.122	.122	.000	89	1.000
Pair 5	RESTOMANTELEXP – RESTOMANTELENC	-.444	1.219	.128	-.700	-.189	-3.459	89	.001
Pair 6	GETRELVENEXP – GETRELVENENC	.056	1.032	.109	-.161	.272	.511	89	.611
Pair 7	SECBREACHEXP – SECBREACHENC	-.278	.561	.059	-.395	-.160	-4.694	89	.000
Pair 8	RESTTONLBANEXP – RESTTONLBANENC	-.056	.976	.103	-.260	.149	-.540	89	.590
Pair 9	NONACEORURPEXP – NONACEORURPENC	.000	1.060	.112	-.222	.222	.000	89	1.000
Pair 10	NONCAPOFDATEXP – NONCAPOFDATENC	-.333	.821	.087	-.505	-.161	-3.851	89	.000
Pair 11	MAKBUZKNTCUSTEXP – MAKBUZKNTCUSTENC	-.722	1.290	.136	-.992	-.452	-5.312	89	.000
Pair 12	UNRELOFINTPROEXP – UNRELOFINTPROENC	.222	1.089	.115	-.006	.450	1.936	89	.056
Pair 13	LAKOFPROBUDGTEXP – LAKOFPROBUDGTENC	-.056	.625	.066	-.186	.075	-.844	89	.401
Pair 14	POWAFAILEXP – POWAFAILENC	.111	1.293	.136	-.160	.382	.815	89	.417

Computer Printout of Researchers' Survey

Discussion of Empirical Results

Table 1 and table 2 explain the expected/encountered benefits of online banking in Nigeria. The tables i.e. 1 and 2 shows a significant difference in the scores for: Competitive advantage expected (COMPADVEXP) (M=2.94, S.D=.916) and competitive advantage encountered (M=3.22, S.D=.715) conditions; t(89), p(.001). Expansion in customer base expected (EXCUSTBASE) (M=3.39, S.D=.682) and expansion in customer base encountered ((M=3.56, S.D=.500) conditions; t(89), p(.025). Loyalty anticipated from customers (LOYALEXP) (M=3.28, S.D=.561) and Loyalty from customer encountered (LOYALTYENC) (M=3.06, S.D=.407) conditions; t(89), p(.001). Expected waiting time reduction (M=3.39, S.D=

.831) and waiting time reduction encountered (M=2.94, S.D=.625) conditions; t(89), p(.000).

However, there is no significant difference in the mean score for: increase in turnover expected (INTURNEXP) (M=3.67, S.D=.474) and increase in turnover encountered (INCTURNENC) (M=3.56, S.D=.500) conditions; t(89), p(.068). Anticipated increase in bank branches network (INCNTWOKEXP) (M=2.89, S.D=.880) and encountered increase in bank branches network (INCNTWOKENC) (M=3.11, S.D=.941) conditions; t(89), p(.114). Positive image anticipated (POSIMAGEEXP) (M=3.39, S.D=.594) and positive image encountered (M=3.28, S.D=.561) conditions; t(89), p(.158). Rise in understanding of activities by employees expected (BETTUNDEXP) (M=3.11, S.D=.880) and rise in understanding of employees encountered

(BETTUNDENC) (M=3.11, S.D=.941) conditions; $t(89)$, $p(1.00)$. Efficient increase in business activities expected (INCBUSACTEXP) (M=3.5, S.D=.503) and efficient increase in business activities encountered (INCBUSACTENC) (M=3.39, S.D=.594) conditions; $t(89)$, $p(.114)$. Reduction in operation cost expected (OPCOSTREDEXP) (M=3.06, S.D=.853) and reduction in operation cost encountered (OPCOSTREDENC) (M=2.94, S.D=.784) conditions; $t(89)$, $p(.401)$.

From the descriptive statistics result in table 1 and 2 above, it was observed that there is a significant difference in the scores of the mean of anticipated benefits and the encountered benefits. Therefore, we can conclude that the differences between condition means are not due to chance. The meaning of this that there was a difference between the benefits expected that online banking will provide for the sector and what is being experienced presently in the sector as regards banks competitive advantage, expansion in customer base, customer loyalty and waiting time in the banking hall. There was however no significant difference in term of efficient business activities, reduction in the operation cost, increase in turnover, increase in branch network, the bank positive image and better understanding of bank activities.

A paired-samples t-test was conducted in table 2 above to compare: the increase in turnover expected (INTURNEXP) and increase in turnover encountered (INCTURNENC); $p(.068)$. Competitive advantage expected (COMPADVEXP) and competitive advantage encountered; $p(.001)$. Expansion in customer base expected (EXCUSTBASE) and expansion in customer base encountered; $p(.025)$. Loyalty anticipated from customers (LOYALEXP) and Loyalty from customer encountered (LOYALTYENC) $p(.001)$. Expected waiting time reduction and waiting time reduction encountered $p(.000)$.

These values are less than .05 level of significance. We then conclude that there is a statistically significant difference between the means of the expected benefits and the anticipated benefits. Therefore, we accept the alternative hypothesis which says there is statistically significant difference between the anticipated and encountered benefits of online banking in the Nigerian Banking system.

A paired-samples t-test was also conducted in table 2 above to compare: the anticipated increase in bank branches network (INCNTWOKEXP) and encountered increase in bank branches network (INCNTWOKENC); $p(.114)$. Positive image anticipated (POSIMAGEEXP) and positive image encountered (POSIMAGEENC) $p(.158)$. Rise in understanding of activities by employees expected

(BETTUNDEXP) and rise in understanding of employees encountered (BETTUNDENC) $p(1.00)$. Efficient increase in business activities expected (INCBUSACTEXP) and efficient increase in business activities encountered (INCBUSACTENC); $p(.114)$. Reduction in operation cost expected (OPCOSTREDEXP) and reduction in operation cost encountered (OPCOSTREDENC) $p(.401)$.

These values are greater than .05 level of significance; as a result of this we conclude that there is no statistically significant difference between the means of the expected benefits and the anticipated benefits. Therefore, we accept the null hypothesis which says there is no statistically significant difference between the anticipated and encountered benefits of online banking in the Nigerian Banking system

Table 3 and Table 4 explain the expected/encountered challenges of online banking in Nigeria. Table 3 showed a significant difference in their scores: anticipated high technology cost (HCOSTECHEXP) (M=3.33, S.D=.750) and high technology cost encountered (HCOSTECHENC) (M=3.61, S.D=.490) conditions; $t(89)$, $p(.003)$. Lack of e-commerce knowledge expected (LECOMMKWNEXP) (M=2.83, S.D=.503) and lack of e-commerce encountered (LECOMMKWNEHC) (M=3.06, S.D=.709) conditions; $t(89)$, $p(.009)$. Anticipated fear of resisting change from manual to electronic (RESTOMANTELEXP) (M=2.83, S.D=.963) and fear of resisting change from manual to electronic encountered (RESTOMANTELEHC) (M=3.28, S.D=.561) conditions; $t(89)$, $p(.001)$. Fear of security breach expected (SECBREACHEXP) (M=2.94, S.D=.527) and fear of security breach encountered (SECBREACHENC) (M=3.22, S.D=.632) conditions; $t(89)$, $p(.000)$. Customer data not captured on the site expected (NONCAPOFDATEXP) (M=2.44, S.D=.689) and customer data not captured on the site encountered (NONCAPOFDATENC) (M=2.78, S.D=.858) conditions; $t(89)$, $p(.000)$. Making business known to customers expected (MAKBUZKNTCUSTEXP) (M=2.50, S.D=.963) and making business known to customers encountered (MAKBUZKNTCUSTENC) (M=3.22, S.D=.715) conditions; $t(89)$, $p(.000)$.

However the following variables shows no significant difference in their scores: Constraint of budgeting expected (LAKOFPROBUDGTEXP) (M=3.11, S.D=.570) and Constraint of budgeting encountered (LAKOFPROBUDGTENC) (M=3.17, S.D=.691) conditions; $t(89)$, $p(.401)$. Effect with power failure (POWAFIEXP) (M=3.06, S.D=.625) and effect with power failure (POWAFIENC) (M=2.94, S.D=1.085)

conditions; $t(89)$, $p(.417)$. Cost of acquiring IT skill expected (ITSKCOSTEXP) ($M=2.83$, $S.D=.963$) and cost of acquiring IT skill encountered (ITSKCOSTENC) ($M=2.83$, $S.D=.838$) conditions; $t(89)$, $p(1.000)$. Lack of e-commerce infrastructure expected (ECOMMINFRCHEXP) ($M=2.83$, $S.D=.503$) and Lack of e-commerce infrastructure encountered ((ECOMMINFRCHENC) ($M=2.83$, $S.D=.604$) conditions; $t(89)$, $p(1.000)$. Getting reliable vendor expected (GETRELVENEXP) ($M=3.11$, $S.D=.570$) and getting reliable vendor expected (GETRELVENENC) ($M=3.06$, $S.D=.709$) conditions; $t(89)$, $p(.611)$. Resistance to online banking expected (RESTONLBANEXP) ($M=2.56$, $S.D=.602$) and Resistance to online banking encountered (RESTONLBANENC) ($M=2.61$, $S.D=.831$) conditions; $t(89)$, $p(.590)$. Unreliability of internet providers (UNRELOFINTPROEXP) ($M=3.11$, $S.D=.661$) and unreliability of internet providers (UNRELOFINTPROENC) ($M=2.89$, $S.D=.570$) conditions; $t(89)$, $p(.056)$. Non accessibility of rural people expected (NONACEORURPEXP) ($M=2.94$, $S.D=.709$) and non accessibility of rural people encountered (NONACEORURPENC) ($M=2.94$, $S.D=.625$) conditions; $t(89)$, $p(1.000)$.

A paired-samples t-test was also conducted in table 4 above to compare: anticipated high technology cost (HCOSTECHEXP) and high technology cost encountered (HCOSTECHENC) $p(.003)$. Lack of e-commerce knowledge expected (LECOMMKNWEXP) and lack of e-commerce encountered (LECOMMKNWENC) $p(.009)$. Anticipated fear of resisting change from manual to electronic (RESTOMANTELEXP) and fear of resisting change from manual to electronic encountered (RESTOMANTELENC) $p(.001)$. Fear of security breach expected (SECBREACHEXP) and fear of security breach encountered (SECBREACHENC) $p(.000)$. Customer data not captured on the site expected (NONCAPOFDATEXP) and customer data not captured on the site expected (NONCAPOFDATENC) $p(.000)$. Making business known to customers expected (MAKBUZKNTCUSTEXP) and making business known to customers encountered (MAKBUZKNTCUSTENC) $p(.000)$.

These values are less than .05 level of significance, we can therefore conclude that there is statistically significant difference between the means of the expected challenges and the encountered challenges. Therefore, we accept the alternative hypothesis which says there is no statistically significant difference between the anticipated and encountered challenges of online banking in the Nigerian Banking system

However the following variables in Table 4 showed no significant difference in their scores: Constraint of budgeting expected (LAKOFPROBUDGTEXP) and Constraint of budgeting expected (LAKOFPROBUDGTENC) $p(.401)$. Effect with power failure (POWAFaILEXP) and effect with power failure (POWAFaILENC) $p(.417)$. Cost of acquiring IT skill expected (ITSKCOSTEXP) and cost of acquiring IT skill encountered (ITSKCOSTENC) $p(1.000)$. Lack of e-commerce infrastructure expected (ECOMMINFRCHEXP) and Lack of e-commerce infrastructure encountered ((ECOMMINFRCHENC) $p(1.000)$. Getting reliable vendor expected (GETRELVENEXP) and getting reliable vendor expected (GETRELVENENC) $p(.611)$. Resistance to online banking expected (RESTONLBANEXP) and Resistance to online banking encountered (RESTONLBANENC) $p(.590)$. Unreliability of internet providers (UNRELOFINTPROEXP) and unreliability of internet providers (UNRELOFINTPROENC) $p(.056)$. Non accessibility of rural people expected (NONACEORURPEXP) and non accessibility of rural people encountered (NONACEORURPENC) $p(1.000)$.

These values are greater than .05 level of significance, we can therefore conclude that there is no statistically significant difference between the means of the expected challenges and the anticipated challenges. Therefore, we accept the alternative hypothesis which says there is statistically significant difference between the anticipated and encountered challenges of online banking in the Nigerian Banking system.

Implication of Findings

Even though there is a positive correlation between the variables under this category, table 1 above showed a high hope of benefits expected of the online banking are not been experienced as thought before the advent of online banking. First the high expectation could be due to what online banking has brought to bear in the developed nation where it has recorded huge success. This might have informed the decision of the various players in the sector not minding other impending factor(s) peculiar to the country. Secondly, the old nature of doing business in Nigeria for example the waiting time in the banking hall has not reduced to the level desired compare to the developed world and the level of computer literacy in the country could also be responsible.

The result in table 2 above showed no significant different among the variables tested. For example, the increase in the network of bank expected before online

banking has been like that till date. The banks studied in this work for example, have increased greatly in term of sizes due to the introduction of online banking. The reduction in the operating cost could be linked to the deployment of chain of computers. This has cut down manual activities where the service of a human is needed and also, the efficient increase in business activities could also be linked to various training and re-training being carried out time without number by various banks for their staffs. This has also helped a great deal.

Conclusion and Recommendation

In spite of the benefit of Electronic Commerce, Automation and Online Banking such as increase in bank branches network, reduction in operation cost among others; the paper identified lack of e-commerce knowledge, fear of resistance to change from manual to electronic, security breach, non-capturing of customer data in total, unreliability of internet providers, making business known to customers through the net and accessibility of the rural people to online banking and power failure as obstacles to full implementation of electronic commerce, automation and online banking in Nigeria. To tap into the full benefits and potential of electronic banking and e-commerce, the operators and the business at large must understand and be abreast with the benefits and challenges of electronic trading.

Recommendations

Notwithstanding the challenges highlighted above, we proffer the following recommendations:

- i. The largest room of improvement is change. Therefore operators should impress it upon employees the need to be responsive to online banking.
- ii. Constant training of employees both local and international on new development in online trading.
- iii. Workshops should be organized for customers periodically on how to keep their data secret especially on how to combine numbers to form password. They should also be vigilant about who stand next to them when using the ATM.
- iv. Even though it is impossible for now considering the dearth of infrastructural facilities in Nigeria, a base could be provided where neighboring villages can access online banking if it is impossible to situate internet facility in each village in the country.

References

- Aarma, August; Vensel, Vello (2001): "Banks' Retail Customer Satisfaction and Development of Bank-Customer relationships," Estonia on the Threshold of the European Union: Financial Sector and Enterprise Restructuring in the Changing Economic Environment, Tallinn: Tallinn Technical University, 2001, pp. 85-106.
- Adesina A. A & Ayo C.K. (2010): An Empirical Investigation of the Level of Users' Acceptance of E-Banking in Nigeria Journal of Internet Banking and Commerce, Vol. 15, No.1
- Adeyinka Adeyemi, Adoption of E-Banking Service Arising in Nigeria, Trade Invest Nigeria.
- Abeyesekera, A., Criscuolo, C., Barreto, E., Gallagher, P. et al., (1999): "Partners Speak Out; Views on Ecommerce", International Trade Forum, Geneva, Volume 2, pp. 23 – 25.
- Alter, S., (1999): "Shopping.com: When E-commerce Isn't a Bargain", Communications of the Association for Information Systems, Volume 2, Article 22
- Alexander, R., (1998): "E-commerce Security: An Alternative Business Model", Journal of Retail Banking Services; New York; Winter
- Amrit B., (2007): "Prospects and Challenges of E-banking in Nepal", The Journal of Nepalese Business Studies. Vol. IV No. 1 2007
- Amadi V.L., (2005): "An Investigation into the Role of Private Sector in Nigeria Higher Institution: A Study of the University of Abuja", International Journal of Research in Education. Vol. 2 No 1&2
- Ayo, C. K., Ekong O. U, Fatudimu, I.T., and Adebisi A. A. (2007): "M-Commerce Implementation in Nigeria: Trends and Issues", Journal of Internet Banking and Commerce, Vol. 12, No.2.
- Bandura A. (1982): "Self-Efficacy Mechanism in Human Agency", Journal of American
- BankAway (2001): "Net Banking Benefits! Sheer Acceleration. Electronic Banking: The Ultimate Guide to Business and Technology of Online Banking", Edited by SCN Education B.V.
- Burnham, B.(1996) "The Internet's Impact on Retail Banking." Booz-Allen Hamilton Third Quarter, (<http://www.strategy-business.com/briefs/96301/>).
- Burr, W. (1996): Wie Informationstechnik Die BankorganisationverändernKonnte, Bank und Markt, Vol. 11, pp. 28-31.

- Coulson, A., (1999), "Electronic Commerce: the Ever Evolving Online Marketplace", IEEE Communications, Volume 37, Issue 9, pp. 58 – 61
- Crane, D.B. and Bodie, Z. (1996): "Form follows functions: the transformation of banking", Harvard Business Review, 74, 2, 109-117.
- Dan, A., Dias, D.M., Kearney, R., Lau, T.C., et al,(2001), "Business-to-business Integration with TPAML and a Business-to-business Protocol Framework", IBM Systems Journal; Armonk, Volume 40, Issue 1, pp. 68 – 90.
- De Souza, N. and I. Von Wiese (2000). "European Union: EU Adopts Directive On e-commerce." International Financial Law Review 19(7): pp. 61-62
- Daniel, E. (1999): "Provision of electronic banking in the UK and the Republic of Ireland", International Journal of bank marketing, Vol. 17, No. 2
- Duclaux, D.(1996): "The Call of the Web," ABA Banking Journal, pp.20-22.
- Echikson, W. (April, 2001): "The Dynamo of E-Banking", BusinessWeek Online.
- Feeny, D. (2000). "The CEO and the CIO in the Information Age". Moving to E Business. L. Wilcocks and C. Sauer. London, Random House: pp. 168-187.
- Ferguson, M., (1999), "Revenue Collection via the Internet: Transaction Cost Components", Collector 1999 proceedings.
- Fraser, J., N. Fraser and F. McDonald (2000). "The Strategic Challenge of Electronic Commerce." Supply Chain Management: An International Journal 5(1): pp. 7-14.
- Grover, V., and Ramanlal, P., (2000), "Playing the E-commerce Game", Business and Economics Review; Columbia; Oct-Dec.
- Gur_u, Catalin.(2002): "E-banking in Transition Economies: the Case of Romania," Journal of Financial Services Marketing, pp. 362 -379.
- Hoffman,D.,Novak,T.andPerlata,M.(1999):"Building Consumer Trust Online", Communications of the ACM, April, Volume 42, issue 4, pp.80 – 85.
- Joze K., Julie F., & Angela S., (2002): "Electronic Commerce Benefits, Challenges and Success Factors in the Australian Banking and Finance Industry", Retrieved on 3/1/2011 from citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.
- Kalakota, R., Oliva, R., Donath, B., (1999), "Move Over, E-commerce", Marketing Management, Chicago, Volume 8, issue 3, pp. 23 – 32.
- Kent, R., Lee, M., 1999, "Using the Internet for Market Research: A Study of Private Trading on the Internet", Journal of Market Research Society, London, Volume 41, Issue 4. pp. 377 –342.
- Koved, L., Nadalin, A., Nagaratnam, N., Pistoia, M., and Shrader, T., (2001): "Security Challenges for EnterpriseJava in an E-business environment", IBM Systems Journal, Armonk, Volume 40, Issue 1, pp. 130 – 152.
- Krejcie, R. V., & Morgan, D. W. (1970): "Determining Sample Size for Research Activities", Educational and Psychological Measurement, 30, 607-610
- Lawrence, E., Corbitt, B., Tidwell, A., Fisher, J., and Lawrence, J., (1998), "Internet Commerce: Digital Models for Business", John Wiley & Sons, Brisbane.
- Leow, Hock, Bee. (1999), New Distribution Channels in Banking Services, Bankers Journal Malaysia, No. 110.
- Lee, C.-S. (2001): "An Analytical Framework for Evaluating e-commerce Business Models and Strategies", Internet Research: Electronic Networking Applications and Policy. pp. 349-359.
- Margaret T., &Thompson S. H. T. (2000) Factors Influencing the Adoption of Internet Banking: Journal of the Association for Information systems. Volume 1, Article 5
- Mohammad M.R., (2009) E-Banking in Bangladesh: Some Policy Implications
- Mols, K. (1998): "The Behaviour Consequences of PC Banking, International Journal of bank Marketing. Vol. 16, No. 5, pp. 195-201.
- Muniruddeen L. (2007). An Examination of Individual's Perceived Security and Privacy of the Internet in Malaysia and the Influence of This on Their Intention to Use E-Commerce: Using An Extension of the Technology Acceptance Model. Journal of Internet Banking and Commerce, vol. 12, no.3
- Mbarika, V.W.A., Okoli, C., Byrd, T.A. and Datta, P. (2004) The Neglected Continent of IS Research: A Research Agenda for Sub-Saharan Africa, Journal of the Association for Information Systems, 6, 5, 130-170

- Nathan, L. (1999), *Www.your-community-bank.com: community banks are going online*. Community and Banking, Federal Reserve Bank of Boston.
- NIGERIA DIRECT: Banking. 10 October 2006 www.nigeria.gov.ng
- Onwuka, E.C. (2006) Another Look at the Impact of Nigeria's Growing population on the Country's Development, *African Population Studies/ Etude de la Population Africaine*, 21, 1, 1-18
- Pohjola, M. (2002) "The New Economy: Facts, Impacts and Policies," *Information Economics and Policy*, 14, pp. 133-144.
- Rahul, D., M. Biju and D. Abraham (2001). "Critical constructs for analysing e-business: investment, user experience and revenue models." *Logistics Information Management* 14(1/2): pp. 137 - 148.
- Riggins, F., 1999, "A framework for Identifying Web-based Electronic Commerce Opportunities", *Journal of organisational computing and electronic commerce*, Volume 9, Issue 4, pp. 297 - 310.
- Salnoske, K., (1997), "Building Trust in Electronic Commerce", *Credit World*, Vol. 85 No 6. pp. 9-11.
- Schwartz, E., (1999,) "Digital Darwinism: 7 Breakthrough Business Strategies for Surviving in the Cutthroat Web Economy", Broadway Books, New York.
- Turban, E., Lee, J., King, D., Chung, M.,(2000), "Electronic Commerce: A Managerial Perspective", Prentice Hall International, Inc. New Jersey.
- Watson, R., Berthon, P., Pitt, L. and Zinhan, G., 1999, "Electronic Commerce : A Strategic Perspective", The Dryden Press, USA
- Whinston, A., Stabl, D. and Soon-Yong, C., 1997, "The Economics of Electronic Commerce", Macmillan Technical Publishing, Indianapolis.
- Winner, L., (1997), "The Neverhood of Internet Commerce", *MIT's Technology review*, Volume 100 No 6 p. 31.
- Sathye, M. (1999), Adoption of Internet banking by Australian customers, *International Journal of bank marketing*, Vol. 17, No. 7, pp. 324-334.
- Schaechter, A., (2002) Challenges of the "E-banking Revolution. *Finance & Development*
- Sannes, R. (2001) Self-service banking: value creation models and information exchange, *Informing Science*, 4, 4, 226-2

A Survey on the Assessment of Iranian External Auditors' Recognition of Materiality Levels

Azar Rahdarian(M.A.)

[Email: rahdarian@gmail.com](mailto:rahdarian@gmail.com)

Hamed hamedian(M.A.)

Tax officer(auditor)

State Tax Organization of Eslamic Republic of Iran Khouzestan Tax Directorate

[Email: h.hamedian@gmail.com](mailto:h.hamedian@gmail.com)

Abstract

The financial statements users always rely on auditors' report in their decision making. Thus, due to the importance of auditors' accreditation, the auditing standards require auditors to design their audit in such a way to gain reasonable confidence from the detection of errors with material effect on financial statements. In this respect, we decided to test the external auditors' familiarity with the qualitative and quantitative materiality levels. The present survey was made about the members of Iranian Association of Chartered Certified Accountants. Firstly the auditors employed in the organization were tested versus the colleague auditors in private institutions. Then, in next assumptions the different job categories of auditors were compared and their familiarity with qualitative and quantitative materiality levels were assessed. The research results indicated that there is an appropriate general consensus among the auditors in the organization and the colleague auditors in private institutions. **Key words:** Quantitative materiality levels; 2. Qualitative materiality levels; 3. Familiarity; 4. Materiality.

The "Materiality" was adopted in auditing and accounting literature since early 1930s. The primary researches may refer to as thresholding, because at the time the researches were to find the materiality threshold or border. This lasted until the mid 1970s. But, the researchers' general tendency about materiality judgments led to the assessment of materiality impact on the audited financial statements users' decision and, thereafter, it was demonstrated that the materiality judgments were qualitative so that the determination of an absolute value materiality threshold became nonsense.

In the next researches on materiality, mainly two tendencies are noticed: firstly, investigation on mathematical and non mathematical models of materiality judgment in audits which were generally unable to present a comprehensive model and, secondly, measuring auditors' general consensus in materiality judgments. (Zarei, 1996)

Subsequent to Holstrum and Messier (1982) little research on materiality was published in the mid 1980. There was renewed interest in materiality in the late 1980s so that it can be said firmly that seldom a topic has been discussed to this extent in accounting and auditing. But,

despite of its wide application in all categories of accounting and auditing this concept is still ambiguity.

Definition of Materiality in Professional Standards

Financial Accounting Standard Board (FASB) and Statement on Auditing Standards (SAS) No. 47 of the American Institute of Certified Public Accountants (AICPA) defines materiality as follows: "Information is material if its omission or misstatement could influence the judgment and decision making of a reasonable person using financial statements."

Section 350 of the Iranian Auditing Standards also gives a definition similar to that submitted by the Financial Accounting Standard Board (FASB).

FASB definition relies on whether the magnitude of the item is probable in affecting the users' judgment while the IASB definition uses the term "could" influence the users' judgment. U.S. standard setters have expressed concern that the use of the word "could" without a modifier such as "reasonably", establishes a very low threshold

of materiality. In light of the fact that the word "could" implies a potentially endless number of possibilities. If the auditor is to design an audit to detect what "could" influence the economic decisions of users, this may impose an unreasonable and impractical instruction to the auditor, especially when dealing with the needs of users of general purpose financial statements.

Through careful reading of definition by the FASB the problem with which the auditors encounter for materiality compliance becomes clear. In this definition it is emphasized on reasonable user and a person who relies on financial statement in decision making. But, in practice, the auditor may be fully unaware of the persons using financial statement as well as the type of decisions taken on the basis of financial statements.

In determining materiality, the FASB and Securities and Exchange Commission (SEC) and, also, the American Institute of Certified Public Accountants (AICPA) have all applied traditional approach (user approach).

Holstrum and Messier believe that there are three major problems in "user approach" to materiality:

1. We know little about how financial statement information is applied by users in investment and decision making.
2. Preparers, auditors and users have dissimilar views of materiality due to their different incentives.
3. There is little information on how materiality judgments made by preparers and auditors affect users' decisions (Messier:2005).

Materiality Concept in Accounting and Auditing

Some accountants and auditors believe that there are two kinds of materiality. Such incorrect belief has even transmitted to the professional literature. For instance, the accountants' international studies group has published the results of its researches under the title of "Materiality Concept in Accounting". While, the results of researches conducted by the Canadian Institute of Chartered Accountants (CICA) has been published under the title of "Materiality in Auditing".

The research group of "Financial Reporting for Non-Profit Organization" has defined the materiality concept in accounting and auditing in the following way:

"Materiality concept in accounting refers to an amount the errors and disorders exceeding it will seriously affect desirable and fairly financial statements. But, materiality

in auditing is to provide a reasonable assurance indicating that the financial statements do not include material misstatement.

Though professional literature has somehow developed this thought under the topic of accounting/auditing orientation, the dominant viewpoint is that the extent of materiality in accounting and auditing is the same, but the content of its application may differ (Zarei, 1996).

While the auditor's responsibility is to express an opinion on financial statements and, in fact, is to assure them for the relevant users. In such cases the auditors should recognize material information affecting the decision taken by a reasonable user of financial statements enabling him/her to properly express an opinion on financial statements. Therefore, it is said that the first step in identifying material information is the proper recognition of materiality levels.

Materiality Levels

It is a level of materiality at which the likelihood of error or fraud in financial statements could affect the financial statement user's decision making (decision makers).

The auditor takes materiality into consideration in the two following cases:

1. For financial statements as a single configuration
2. For each account balance, transactions group and disclosures

Meanwhile, issues as those mentioned below can also affect the materiality amount:

- Regulatory and Legal Requirements
- Considerations about each account balance as stated in financial statements and the relationship between these items

Regarding that the auditors in all audit steps, from the planning and execution to the reporting of results obtained, are involved in the considerations of materiality. In most cases the auditor encounter conditions under which he cannot make decision on the materiality and/or immateriality of detected misstatements. That's why there is still discussion on the definition of this "size" or the extent required to affect the users' decision. Investigations have shown that the auditors consider the materiality concept as well as its assessment through taking into account some quantitative and qualitative factors. In most cases the auditor encounter such conditions under which he cannot make decision on the materiality and/or immateriality of detected misstatements. In such situations, the auditor should determine the expected probable impact of these misstatements and where the total misstatements he/she identifies go beyond the materiality

border he/she takes them into consideration and examines their impact on financial statements.

Sometimes the aggregation of multiple less material misstatements may become material, or the aggregation of multiple material misstatements may become substantial. Thus, the auditors require guidelines or standards contributing them in identifying material misstatements in the financial statements.

In this regard many attempts have been made by specialized committees and associations and the professional publications to set up professional guidelines concerning materiality the most important of which are discussed here in brief:

The Accountant International Study Group composing of professionals from Canada, UK and USA, after examining the relationship of the concept of materiality with some accounting cases and expressing professional opinions as well as complying with regular rules in their relevant countries, reached the following overall and general conclusions:

1. Materiality is a matter of professional judgment. An item is material if knowing it would affect financial statements users.
2. An amount is not considered material because of its quantity, rather other factors, in addition to what stated below, should be taken into account on making a materiality decision:
 - a) The nature of each item
 - b) The amount itself in association with:
 1. Generality of financial statements
 2. Total amount of accounts to which this item is an element or should be an element
 3. Items related thereto
 4. Similar amount in previous years or amount predicted for future years

In addition, the Australian Auditing Principles Committee in October 1972 proposed to apply an extent equals 5 to 10 percent for making a materiality decision.

Quantitative approaches for making materiality judgment are categorized into one of the four following groups:

Single Rules Method:

These are the "Rules of thumb" using a single financial variable for materiality estimation. The auditors usually provide three or four such rules according to policies and each firm's auditor should select the most appropriate rule from among the available methods. Based on his assessment

of quantitative factors, the auditors would select a rule to be the most appropriate one to estimate the materiality of client's amounts.

Here are the most common examples for single rules as follows:

- a) 5% of pre-tax profit
- b) 1.2% of total assets
- c) 1% of equity
- d) 1.2% of total revenues

Variable Rules:

These are similar to single rules but they vary in that they provide a range of different materiality levels for companies of different sizes. In this method, as with the single rule, the auditor uses the assessment of quantitative factors to decide what materiality level to select to be suited to auditing range.

The possible examples for variable or size rules are as follows:

- a) 2 to 5% of gross profit if it is less than USD 20,000. Dollars
- b) 1 to 2% of gross profit if it is between USD 20,000. -1,000,000.
- c) 1.2 to 1% of gross profit if it is between USD 100,000. – 100,000,000.
- d) 1.2% of gross profit if it is over 100,000,000.

Average or Blend Method:

This method takes 4-5 rules of thumb and, then, it weighs each rule according to rule, in proportion to degree and average (an equal weighting). The blending and averaging process is probably an indirect way to consider quantitative factors. An example of this method is to take five single rules with equal weights:

XX	5% of assets
XX	5% of income
XX	5% of pre-tax profit
XX	2% of gross profit after depreciation
XX	1% of shareholders' equity
XXX÷5	Materiality

Formula Method: This method uses a mathematical formula and the materiality level for a large sample of institutions is determined through statistical analysis. Since the rules of thumb are used the individual materiality levels are taken into consideration. In fact, formula methods are another form of average or blend method. The most

widely known formula method is the one used in KPMG formula although there are other methods and formulas in audit literature. In reference to interpretations, in 1998 the KPMG was as follows:

$$\text{Materiality} = 1.84 (x)^{\frac{2}{3}}$$

where 1.84 and the exponent are constants having been empirically computed through review and statistical sampling of a large number of companies and the variable (x) represent the greater figure obtained from the comparison of assets with revenues. That is, the greater figure (assets or debts) will be taken as variable (x).

Since this method is the only way to meet the variability of materiality judgment compared to two other methods and, in total, the amount obtained from KPMG formula is almost twice as much as other methods and, also, the extent of materiality applied for other auditors is different. It is utilized by most institutions.

Applying the said quantitative guidelines in a trading unit and in different industries would show various materiality levels. This is way the ratio of assets, profit and sales in different industries and companies differ a lot. In addition, each of above guidelines has intrinsic failure.

Overreliance on quantitative materiality thresholds (such as 5% of net income) may cause auditors to waive quantitatively immaterial but qualitatively material audit differences and, consequently, undermining the quality of audited financial reports.

Securities and Exchange Commission's (SEC's) staff accounting bulletin No. 99 particularly reminded that numerical thresholds are not acceptable alone. According to the instructions of the said commission, the management should measure qualitative issues whether such misstatement makes change in revenues or be related with critical and important part of trade.

Both the Securities and Exchange Commission and Financial Accounting Standard Board (FASB) in the United States believe that despite the use of quantitative criteria, the auditors should not simply rely on quantitative factors and figures in their materiality judgment. Fully reliance on quantitative factors in materiality assessment whether in accounting literature and/or legally is not allowed and acceptable.

Also, Financial Accounting Standard Board (FASB) emphasized that materiality is not a qualitative concept. In fact, FASB refused to take a formulatic approach in

determining materiality and has accepted an approach in favor of one that incorporates all related conditions.

Qualitative factors are usually more difficult to assess than the size of the misstatement or omission. Qualitative factors often require subjective judgment and evaluation in light of other information that may not be easily accessible to the auditor during audit. In practice, though speaking frequently about the qualitative factors, it is more likely to ignore these factors.

Securities and Exchange Commission (SEC) has listed qualitative factors that cause little misstatements of a financial statement item as follows:

1. Arises from an item capable of precise measurement or from an estimate, and, if so, the degree of imprecision inherent in the estimate
2. Masks a change in earnings or other trends
3. Hides a failure to meet analysts' consensus expectations for the enterprise
4. Changes a loss into income, or vice versa
5. Concerns a segment or other portion of business that has been identified as playing a significant role in operations or profitability
6. Affects compliance with regulatory requirements
7. Affects compliance with loan covenants or other contractual requirements
8. Increases management's compensation--for example, by satisfying requirements for the award of bonuses or other forms of incentive compensation
9. Involves the concealment of an unlawful transaction
10. In section 1-6 of its new auditing guide the Audit Organization in Iran, as a specialized competent authority to set up terms and principles of accounting, has taken into account the materiality computation under ordinary conditions on the basis of 3% of the total average assets and sale/income which, with increasing basic amount, the materiality level is reduced from 3% to 1%. (Publication 150) Additionally, allocating materiality to accounts (items in the balance-sheet and profit and loss account) should be considered equal 3 times as much as the assessed materiality amount determined in financial statements, with stating that the maximum materiality amount allocated to each account cannot exceed 60% of the assessed materiality. In the guideline of organization, no details and logic underlying a suggested basis have been given for materiality guideline (3% of total average assets and sale) and, also, for the calculation of coefficient 3 or making materiality amount three times more in order

to allocate materiality to accounts balance. But the above basis is not complied with the findings of other researches, in the current case where the assets and the sale are less material than the net profit of materiality. (Kasiri, 2003).

Now knowing that accounting and auditing profession's knowledge about the information needs of a wide range of users on financial reports, on one hand, and the type of decisions they made and, also, the variation of information affecting their decisions and adjusting and/or changing such decision is so rare, we found out why the materiality concept is so difficult and why there is no professional standard either at international or at national level to cause precedent and integration in the performance of this profession members about the assessment of materiality level of the cases given in accounting and auditing and that why some professional institutions have simply taken measures to set quantitative guidelines which are both very limited and still relying on the judgment made by the professionals of this occupation.

But, seemingly, after acceptance of all limitations and ambiguities existing in professional definitions and instructions it would be better to prepare different minimum angles needed by auditors to make judgment and to submit them as a list to the profession members.

Regarding the necessity of existing professional guidelines and instructions, Professor Anderson believes that "materiality judgment with no guideline would lead to opposite results under similar conditions and will waive the profession's general confidence".

And in case of the public's non-confidence in the auditing profession the main role and task of auditors will be questioned!

Despite the profession's feeling strong need to the existence of professional instructions to guide auditors on making reasonable decisions some disagreed with setting guideline and they believe that relying on guideline would lead to taking unaware and non reasonable decisions simply on the basis of guideline and using individual judgment will be abandoned.

But, in practice, making decision about "real materiality" in certain conditions requires a judgment which is not so simple. There is no simple and undefined guideline to enable auditor make a decision about when something is immaterial, material or very material.

In Staff commission bulletin No. 99 describes that: The materiality of a misstatement also depends on that in which part of the financial statement this misstatement has been appeared, and the auditors should take into consideration

not only the size and the magnitude of the misstatement but also the materiality of information in each part of the financial statements as a whole.

Warning Signs of Possible Misstatement

Conditions may indicate the probability of a material misstatement caused by fraud or error in financial statements. The auditor should be conscious of the following situations, submitted as a sample, which may indicate the probability of a fraud or error:

1. Management's imposing non-reasonable temporal deadlines for the auditing completion.
2. Management reluctance into presenting true reports to the competent third persons, i.e. supervisory authorities and banks.
3. Imposing limitation by the management on the scope of audit work.
4. Identifying material issues not been previously disclosed by the management.
5. Major items the auditing of which is difficult.
6. Non-reasonable insisting on applying bold (versus conservative) accounting standards.
7. Opposite or insufficient evidences submitted by managers or staffs.
8. No ordinary written evidences such as manual modifications of documents or preparing manual documents where these documents are usually prepared by electronic machines.
9. Presentation of information with no reasonable delay or reluctantly.
10. Entirely imperfect or insufficient accounting records.
11. Transactions without sufficient or valid evidences.
12. No ordinary transactions in term of nature, size or complexity, especially if occurred at the end of fiscal year.
13. Transactions not recorded in compliance with managers' general or special permit
14. Major differences between general ledgers and general journals or between subjective inventories counting and relevant accounting records having not been reviewed and amended timely and properly.
15. Insufficient control on computer data processing (for example, so many errors or delays in processing).
16. Major differences between real results and expected results obtained from analytical examinations.
17. Receiving replies on confirmations is far less than expected or major conflicts on the basis of received confirmations.

18. Evidences indicating excessive splurging in managers or staffs' life
19. Annual outstanding accounts
20. Overdue accounts and documents receivable

In January 2006 the American Chartered Accountants publication issued an article titled "Materiality Assessment" presenting a modern method called "Fuzzy Logic Approach" in materiality assessment.

This article has summarized the main problem encountering auditors in the materiality estimation in two categories: First, auditors inevitably take binary decisions (material versus immaterial) and, then, they should weigh the qualitative factors to achieve this. This system enables the auditors to assess materiality as a continuous characteristic by allowing a misstatement to possess a degree of materiality between 0 and 1 and to consider important qualitative factors in materiality assessment.

In practice, the auditor should take such decision about materiality as having been too underestimated in audit profession; a binary decision regarding each omission and misstatement, both individually and in aggregate, as to whether the item is material or not.

A fuzzy logic approach would make possible to assess materiality as a continuous characteristic by allowing a misstatement or omission to possess a degree of materiality between 0 and 1.

The theory of fuzzy sets and fuzzy logic was first introduced by Professor Askarzadeh from the University of California, Berkeley. Fuzzy logic is the logic that underlies the paradigm of the partial membership of an item in a set. In fact, rule-based fuzzy and classic expert system can be considered as an advisory board consisting of a large number of advisors. In current application, the auditor – the decision maker – would ask each advisor to review an omission or misstatement and to decide whether it is material or not. Imagine that each advisor, when making decision on materiality, would focus on only one aspect of the omission or misstatement. For instance, one advisor may consider only the size of the given omission or misstatement, while another advisor may think of only the extent to which it increases management allowances.

Therefore, each advisor would take into account one of the rules in the expert system. Finally, the decision maker would assimilate the various opinions of the advisors and come to the conclusion about the materiality of the omission or misstatement. In this way, the decision maker would also incorporate his/her own assessment of the advisors' aptitude in making materiality judgments.

In a classical system, each advisor is obliged to make a binary decision: the omission or misstatement is material or it is not material. The first advisor, for instance, may respond that, based on the size of the omission or misstatement, the omission is not material while the second advisor may say that, as it increases the management allowances, it is material.

In a fuzzy system, each advisor is allowed to announce his/her materiality judgment as a number between zero (0) and one (1); i.e. one (1) to express the judgment that the omission or misstatement is material, and zero (0) to express the judgment that it is not material. The advisors' responses indicate their initial assessments on materiality that will be modified by the advisor later on, as desired.

In a classical system, it is assumed that every rule is valid, i.e. every advisor's viewpoint is taken into account in the materiality assessment. This is while in a fuzzy system, each rule is not considered in the materiality assessment. The decision maker is required to assign a validity value to each rule. When determining such validity value for each rule (which corresponds to the viewpoint of each of the auditors, the advisor would consider his/her assessment of the aptitude of every advisor. For example, the auditor may assign a rather low, 0.35, validity value to the rule: If (Size of Misstatement is Medium) then (Misstatement is Material).

The last step is to assimilate the final assessments made by every rule into a final materiality assessment. The standard approach is to select the maximum final materiality assessment:

Final Materiality Assessment = Max (Final Materiality Assessment as per all Rules)

In the classical system, the misstatement could be assessed as material when one or more of the rules represent that the misstatement was material.

For the purpose of materiality assessment, an effective fuzzy logic expert system should incorporate rules that take into consideration both the quantitative and the qualitative aspects of misstatement. The quantitative aspects should also contain both the size of the misstatement and the precision with which it is measured. The qualitative aspects should include those stated by the Financial Accounting Standards Board.

Upon assigning a fuzzy degree – any number in the interval [0 and 1]- the auditor will be provided with more flexibility and greater precision in materiality assessment and, also, he/she will get deeper insight regarding subsequent testing and investigation.

Through providing a formal model structure, the fuzzy system would formalize and document the materiality assessment process. This model would require the auditors involving in definite assessment of each one of the qualitative and quantitative factors. Also, in this way, the designers would be obliged to state each rule specifically and assign a definite validity to each rule. This might lead to the facilitated communication of the audit team with the client as well as the increased consistency of the process the auditors apply in different years.

To build a valid and reliable fuzzy system for materiality assessment, the auditing experts are obliged to extract the quantitative and qualitative factors that auditors should apply as well as the validities of these fuzzy rules. Upon completing an initial design, a formal feedback system may be incorporated to improve the future performance. If materiality assessments later prove to be incorrect, such system can modify itself to improve future assessments under similar conditions.

The results of empirical researches show that there are various qualitative and quantitative factors affecting the materiality judgments and auditors' disclosure. The Financial Accounting Standard Board (FASB) pointed out that it is helpful to use multivariate models in order to continue researches and to provide required conditions for the continuation of the integrated materiality guidelines is helpful.

The Necessity of Research

As the external audit aims at audit within the auditing standards framework to be reasonably assured of not existing error or "material" misstatement in financial statement and to declare on its "optimal presentation" in compliance with accounting standards from all "material" aspects and, finally, its assurance. Financial statements users rely on the auditors' report in their decisions and in cases where the financial statements are misleading the investors and the assurers incurring financial losses will accuse the auditors of professional improper performance and will legally prosecute them to compensate them for damages they have incurred.

For this reason and due to the importance of auditors' role in assurance, the auditing standards stipulate that the auditors should design their audit in such a way to get reasonable statistical confidence in detecting errors with material effect on financial statements. But, what basis is there for auditors in the recognition of immaterial, material or very material cases? The previous researches would

conventionally benefit from quantitative rule of thumb in their materiality judgments and this excessive relying on the quantitative materiality threshold (e.g. 5% of net income) would restrain auditors from detecting errors with no significant quantitative materiality but with qualitative materiality and would weaken the quality of audited financial reports.

Although occupational standards and instructions are submitted in different countries, there has been still no comprehensive and precise guideline concerning the identification of "material" factors to entirely guide auditors in identifying materiality, and this has been assigned to the auditors' professional judgment.

Thus, it seems that the necessity of auditors' acquaintance about materiality levels in the accounting reporting process is an inevitable essentiality. In this regard we turned out to examine the external auditors' acquaintance about the quantitative and qualitative materiality. We hope that the results of this research will be used by competent organizations through clarifying external auditors' weak and strong points and, in this way, it will be possible to provide auditors' better and more recognition of different materiality levels. And, in conclusion, more accurate judgments and, lastly, financial statements with more reliability will be put at the dispose of financial statements users.

A Review of Prior Researches

Firstly, theses and researches conducted about the materiality in Iran are explained and, then, it will be dealt with studies and researches made outside of Iran.

Studies made in Iran

In a research titled "Concept of materiality in audit reporting process" Zarei (1996) had the following objective:

As the determination of materiality level during audit reporting step is a judgment process and no certain guideline has been published to contribute auditors in determining the materiality degree of items and events is there consensus between Iranian auditors in the determining of the materiality of above cases? Upon using statistical methods their researches came to this conclusion that Iranian auditors with a level of 95% within groups (auditors of private and public sectors) have appropriate consensus regarding the degree of materiality of items and events and appropriate kind of audit reporting, but such consensus do not exist between groups.

Hasas Yeganeh, Kasiri (2003) conducted a research aiming at: (1) Identifying quantitative and qualitative factors affecting the determination of materiality level; (2) Determining the extent of consensus among auditors in selecting and applying quantitative guidelines of materiality in completed audit projects; (3) Recognition of problems, obstacles and the necessity of setting quantitative guideline for materiality in Iran.

This research was made in two steps by using poll (questionnaire) and subjective observation of evidences (data mining). The statistical population of the first step included the audit managers of audit organization and the audit managers of the private sector cooperating with that organization, including certified accountants member of the Association of Chartered Certified Accountants and the trusted auditors of Tehran Securities and Exchange Organization among which 70 persons were questioned in selected sample.

The statistical population of the research's second step included audit files and reports of those firms audited by audit organization within the period from 1999 to 2001 and have received not desired declaration (conditional, failed and no declaration). Of this population 223 firms were selected and examined.

The results of this research showed that:

1. Quantitative factors (determinants) such as the size of judgment item, amount of total assets, average amount of total assets and earnings, amount of net profit, amount of capital owners' equity, main amount and main classes depending on judged item in financial statements are the most important quantitative factors in the determination of the materiality level of audit.
2. Qualitative factors (modifiers) such as universality of error impact on different sections of financial statements, the relationship between the judged item and the transactions of affiliated and doubtful persons, the nature of being deliberate or indeliberate, ordinary or not ordinary, estimable or determinable, violation of the regulations and legal requirements and the probable hazard of auditing the judged item, are among those qualitative factors which may lead to modified level of audit materiality.
3. Quantitative guidelines indicating total assets, total incomes, their average, investors' equity (as 1-3 percent) as well as the quantitative guideline based on pre-tax net profit (above 5 percent) are agreed by audit managers.

Vahidi Elizei, Rahdarian (2008). The research was made aiming at the assessment of factors affecting auditors'

judgment and the determination of the related degree of materiality in detecting the misstatement of financial statements. After theoretical studies, in the present survey the quantitative and qualitative factors are identified and categorized as 28 factors into two groups and dispatched to the external and internal auditors as a questionnaire to which 131 external auditors and 59 internal auditors responded.

The results of this research showed that; despite the existence of overall difference in auditors' views of factors efficiency, the auditors rated them in a different order. Thus, external auditors considered 8 quantitative and 5 qualitative factors as the most effective factors while the internal auditors considered 6 quantitative and only 1 qualitative factors as the most effective factors.

The research results also indicated that the external auditors consider qualitative factors more effective and the internal auditors consider the quantitative ones.

Researches Conducted in Other Countries

Messier (1983) examined the impact of experience, type of company and financial variables on auditors' judgment about disclosure and materiality and concluded that:

1. Net profit is the most significant variable while the incomes procedure occupies the second rank in term of materiality.
2. Auditors' judgment about disclosure and materiality is affected by experience and the type of company.

Estes and Reames (1988) studied the impact of auditors' personal characteristics (such as experience, education, place of employment, frequency of materiality decisions, age and gender) on auditors' materiality judgment. 596 chartered accountants responded to their survey.

Their research results showed that only the age and place of employment significantly affect the participants' judgment.

Chewning et al (1989) examined the difference in materiality decisions for optional accounting changes (LIFO) toward non-optional accounting changes (deficits financing and currency exchange). They found out that the materiality of the change impact on profit is the most significant factor in auditors' materiality judgments. Chewning et al also realized that the report amendments were too high for desirable changes. Finally, there was restricted evidence indicating that eight large firms would less likely change their comments rather than other firms.

In his essay, Z.M. Khalifa (1992) pointed out the results of previous researches on the relative importance of factors affecting materiality judgments, of which the factors associated with profit (i.e. factors of profit and loss statements) have been known as more important than the items of balance-sheets in the formation of final materiality judgments, and considering that the most parts of changes in all studies have been assigned to the impact of judged item's impact on factors related with profit, the judged item impact on the factors of balance-sheet had led to ambiguous results. He concluded that a major problem in previous researches on the review of the materiality of factors effective on judgment was to conduct researches within the framework of a firm facing no difficulty. Thus, in their PhD thesis, they studied the impact of firm's financial conditions on materiality judgments and auditing disclosure. The results of this research showed that during formulating materiality judgments, the auditors tend to the judged item impact on factors related with profit. But, in firms facing financial difficulty the balance-sheet items are converted into material components in materiality judgments. (Barati 1996)

Tuttle et al (2002) examined the appropriateness of the conventional materiality thresholds utilized by auditors from the user's viewpoint in an empirical approach for market. For this purpose, 12 market sessions each attended by six dealers and composing 12 external three-minute commercial courses were held.

72 top students of the Bachelor program of business administration, represented as semi-skilled investors, were given financial information. This information was included in material and immaterial misstatement. In fact, the materiality level was manipulated in two levels: conservative materiality (based on 5% pre-tax profit or 25% from net sales) and high materiality (based on 10% pre-tax profit or 5% from net sales).

The comparison of prices brought by participants in market indicated that the not disclosed misstatements within the framework of materiality thresholds are consistent with the present audit procedures. That is, they had no impact on market prices within or below materiality threshold, but the large misstatements did.

Hypothesis

Considering the objectives of research and determining the external auditors' recognition of materiality levels, the following hypotheses were formed:

1. No significant difference exists between external auditors employed in organization and those in collaborator private institutions in term of the recognition of materiality levels.
2. No significant difference exists between shareholders and managers in recognition of materiality levels.
3. No significant difference exists between shareholders and supervisors in recognition of materiality levels.
4. No significant difference exists between managers and supervisors in recognition of materiality levels.
5. No significant difference exists between supervisors and chief auditors in recognition of materiality levels.
6. Auditors' recognition of materiality levels is related with their experience.
7. Auditors' recognition of materiality levels is related with their education.

Methodology of Research

The present research is a descriptive survey conducted with the aim of the assessment of auditors' recognition of materiality levels through theory concepts testing in the Iranian Association of External auditors by statistical surveying method. Two major methods were used for information collection. Firstly, a rather detailed library studies were made for case explanation and research. During this step, in addition to study on the theoretical fundamentals of research, mainly extracted from accredited scientific articles, publications and quarterly journals, the audit standards and instructions of Iran and other countries were studied carefully and, then, based on these studies, measures were taken to prepare a questionnaire.

This questionnaire was designed in two parts as:

Part 1- General information (demography)

Part 2- Determination of auditors' acquaintance with quantitative and qualitative materiality levels

The preliminary questionnaire was given to the professors of accounting, statistics, methodology of research as well as the chief managers of audit firms specialized in the research topic (as control group outside of the statistical population). They were asked to express their views on the questionnaire content. Thereafter, required amendments were made in the questionnaire as suggested by above group. Then, 3-point Likert scale was used to measure research variables (quantitative and qualitative levels as external variable and the extent of auditors' recognition as dependent variable). Interestingly, John Best also prefers this method to 5-point scale. (Zarei:1996) After the

questionnaires were collected, the responses were ranked on a three-point rating scale (1="Very low", 2="Low" and 3="High"). The reliability of the questionnaire was measured by Cronbach's alpha coefficient and the results were analyzed by SPSS statistical software. A Cronbach's alpha of 0.87 was obtained indicating that the reliability of questionnaire was good. The review of all of the individual questions showed that the Cronbach's alpha coefficient will not increase or decrease significantly when any individual question is omitted and this indicates that the questions had good internal consistency.

Statistical Sample and Population

The statistical population of this research consists of external auditors. But, for the applicability of the research as well as the possibility of generalization and comparison of the obtained results, in the first hypothesis the statistical population was divided into two categories of auditors employed in auditing organization and auditors in collaborator private institutions. Random simple sampling method was used for sample selecting. Statistical formulas were applied to calculate the sample size with using sampling methods. And the simple size was calculated as $n \cong 170$. Therefore, in order to eliminate the modifications occurred on variance due to the selection of maximum and minimum scores based on Likert scale, 200 questionnaires were given to external auditors. And, finally, 131 questionnaires returned by the external auditors were used.

Statistical Methods of Research

SPSS is used to perform analysis of research results. Firstly, descriptive statistics including frequency, frequency percentage, mean and standard deviation was applied for all hypotheses and, then, inferential statistics; T-Test, analysis of variance and the Binomial Test were used in hypotheses 2 to 5 to measure different groups' viewpoint. Also, ANOVA test was used for hypothesis 1 as well as assessing the impact of some variables such as experience and education on the auditors' recognition. The failure rate of hypotheses is $P > 0/05$.

Conclusion on Research Hypothesis

Conclusion on the First Hypothesis

ANOVA test was used to assess the general consensus among external auditors (auditors employed in audit

organization and the auditors employed in collaborator private institutions) in the recognition of materiality levels. The results of this test, as indicated in Tables 1 and 2, show that the sample variances for organization auditors and out-of-organization auditors is less than the 5% level of significance. That is, there is no difference between these groups in the recognition of materiality levels. Such general consensus existing between auditors inside the organization and the auditors in collaborator private institutions would result in taking similar decisions under similar conditions and will resolve most doubts concerning professional judgments.

In a similar research, Zarei (1996) found such consensus among the managers of auditing organization and the managers of collaborate private institutions. Pattilo (1976) also came to this conclusion that the assessments of materiality among auditors, financial managers, financial analysts and the university authorities differ a lot in similar conditions and, in this respect, the financial managers have the highest level consensus on the materiality thresholds.

Table 1 : P-Value Analysis – Hypotheses 1

	Acquaintance with Quantitative Materiality Levels	Mean	Std. Deviation	Sig
Total	High	.333	.000	.000
	Low	.651	.070	
	Very Low	1/000	.000	
	Total	.491	.212	
Auditors Employed in Audit Organization	High	.333	.000	.000
	Low	.667	.000	
	Very Low	1/000	.000	
	Total	.449	.201	
Auditors Employed in Private Institutions	High	.333	.000	.000
	Low	.667	.000	
	Very Low	1/000	.000	
	Total	.478	.321	

Table 2 : P-Value Analysis – Hypotheses 1

	Acquaintance with Quantitative Materiality Levels	Mean	Std. Deviation	Sig
Total	High	.342	.054	.000
	Low	.667	.000	
	Very Low	1/000	.000	
	Total	.519	.227	

	Acquaintance with Quantitative Materiality Levels	Mean	Std. Deviation	Sig
Auditors Employed in Audit Organization	High	./333	./000	.000
	Low	./667	./000	
	Very Low	1/000	./000	
	Total	./478	./218	
Auditors Employed in Private Institutions	High	./348	./070	.000
	Low	./667	./000	
	Very Low	1/000	./000	
	Total	./541	./229	

Conclusion on the Second Hypothesis

The results of statistical analysis and the research hypotheses testing shown in Table 3 indicates that the significance level of the sample variances for shareholders and managers is 0.33 which is greater than the 5% level of significance and rejects the hypothesis at 95% confidence level. In other words, the shareholders and managers do not have similar recognition of materiality levels. The examination of the mean values calculated for the two groups show that the shareholders have higher recognition of materiality levels than the managers. But, among the quantitative and qualitative materiality levels, the means of the qualitative factors have higher rank indicating that the both groups have more recognition of qualitative materiality levels than quantitative materiality levels.

Conclusion on the Third Hypothesis

The results of the single-sample T-test for the two groups of shareholders and supervisors, as depicted in Table 3, show that there is appropriate general consensus between these two groups regarding the recognition of materiality levels. But, when the materiality levels were examined separately, the level of significance of the sample variances for the quantitative materiality levels was less than 5%. But, this hypothesis was rejected in qualitative materiality levels at 95% confidence level. That is, the shareholders and the supervisors have similar recognition of quantitative materiality levels but they have no similar recognition in qualitative materiality levels. The comparison of the mean

values of these two groups show that the supervisors have a better recognition of materiality levels and, in this respect, the qualitative materiality levels in both groups have higher mean values.

Conclusion on the Fourth Hypothesis

As shown in Table 3, a one-way analysis of variance (ANOVA) for the two groups of managers and supervisors indicate the confirmation of above hypothesis with a 0.001 significance level. In fact, the managers and the supervisors have appropriate general consensus on the recognition of quantitative and qualitative materiality levels. Of course, the comparison of the means indicate that the supervisors have more recognition of materiality levels as compared with supervisors.

It should be mentioned that in two groups above, as the same as the two previous groups, the qualitative materiality levels have higher mean values than the quantitative materiality levels.

Conclusion on the Fifth Hypothesis

The results of T-test for the two groups of supervisors and chief auditors, as illustrated in Table 3, show that the level of significance of the sample variances is 0.381 indicating that no general consensus exists between supervisors and chief auditors on the recognition of materiality levels. For assessing these two groups' level of the recognition of materiality levels, the means of these two groups were compared and it was determined that the chief auditors have higher recognition of materiality levels as compared to the supervisors. Also, it was noted that the supervisors and the chief auditors also recognize qualitative materiality levels more than quantitative levels.

In the summary of hypotheses 1-5 it can be said that the external auditors, in general, identify qualitative materiality levels better than quantitative materiality levels. But, there is no appropriate general consensus among groups so that the means of the considered sample showed that different job classes of auditors identify materiality levels in the following order of priority:

Chief auditors>supervisor>partner>manager

Table 3: T-test Analysis

	Materiality Levels	Position	Mean	Std. Deviation	p-value
H2	Quantitative	Partner	./386	0/124	0/293
		Senior Manager	./367	0/102	
	Qualitative	Partner	./438	0/194	0/33
		Senior Manager	./383	0/122	
	Total	Partner	1/236	0/386	0/115
		Senior Manager	1/125	0/319	
H3	Quantitative	Partner	./386	0/124	0/000
		Manager	./523	0/190	
	Qualitative	Partner	./438	0/194	0/103
		Manager	./559	0/240	
	Total	Partner	1/236	0/386	0/008
		Manager	1/625	0/602	
H4	Quantitative	Senior Manager	./367	0/102	0/000
		Manager	./523	0/190	
	Qualitative	Senior Manager	./383	0/122	0/000
		Manager	./559	0/240	
	Total	Senior Manager	1/125	0/319	0/000
		Manager	1/625	0/602	
H5	Quantitative	Manager	./523	0/190	0/210
		Supervisor	./600	0/255	
	Qualitative	Manager	./559	0/240	0/926
		Supervisor	./600	0/255	
	Total	Manager	1/625	0/602	0/249
		Supervisor	1/800	0/767	

Conclusion on the Sixth and the Seventh Hypotheses

Success due to correct judgment, correct judgment due to experience and experience arisen from error in judgment is ever obtained over time. Therefore, it is expected that broad experience and academic education are directly related with auditors' recognition of materiality levels and, finally, would result in more correct judgment on financial statements. Thus, we tested the relationship of these two variables with the recognition of materiality levels. The tables of frequency distribution show that 23.7 of auditors participating in the test had more than 20 years of experience in auditing. ANOVA test was used

to generalize the results of the considered samples to the whole statistical population. The results showed that (Table 4) experience was a significant factor in the recognition of materiality levels while there was no significant relation between the auditor's education and his/her recognition of materiality levels (Table 5). Therefore, in compliance with other similar researches such as Messier (1983), Krogstad (1984), Carpenter and Dirsmith (1992), the present survey proved that broad experience would always lead to more correct judgments and more valid financial statements in auditing profession. But, this was not confirmed by research conducted by Estes and Reames (1988).

Considering that 77.9% of participants in the test had graduated in accounting; 30.5% held Master's degree and 61.8% Bachelor's degree; unfortunately the non confirmation of the said hypothesis indicates that the academic degree had little effect on auditors' more recognition of materiality levels.

Table 4 : T-test Analysis

	Experience	Mean	Std. Deviation	sig
H6	... <2 Years	2/000	./620	./000
	2-5 Years	1/518	./612	
	5-10 Years	1/740	./655	
	10-15Years	1/416	./600	
	15-20 Years	1/384	./545	
>20 Years	1/193	./441	
	Total	1/511	./620	

Table 5 : T-test Analysis

	College Degree	Mean	Std. Deviation	sig
H7	Diploma	1/166	./288	./069
	Upper Diploma	1/000	./000	
	Bachelor Degree	1/635	./637	
	Master of Science	1/325	./525	
	Ph.D	1/416	./801	
	Total	1/515	./619	

Suggestions

During research, with considering information resources associated with research topic as well as the results and consequences of the present research, some suggestions

were taken into consideration which are classified into two categories. The first category includes suggestions related with research topic and the relevant findings and, the second are the suggestions for future researches.

Suggestions from Research Findings

Revision of Accounting

As the auditors' recognition of materiality levels is directly related with the kind of auditors' declaration (acceptable, conditional and rejected), the auditors' correct judgments about material information is an important element to give users confidence to general reporting. The option can be correct or incorrect; benefiting from others' experiences may contribute the correct option. The findings of the present research once again showed that the auditors' judgment and declaration is a process directly related with auditors' experience so that the experienced auditors would recognize materiality levels better than less experienced auditors. But, despite the researcher's prediction, the increase in education level had no impact on auditors' more recognition of materiality levels. Thus, it seems that it is an inevitable necessity to put more focus on the issue of materiality in the courses of Bachelor's and Master's program.

Since in Iran there are currently Bachelor's, Master's and PhD programs in accounting and, also, the auditing, which is itself a broad and deep knowledge, is simply taught as a credit course of accounting; 6 credits in Bachelor's and 2 credits in Master's program. Unfortunately, the research results show that auditing could not provide required efficiency for auditors.

Thus, the researcher suggests the fulfillment of the following cases in order to achieve the education level's efficiency and effectiveness on the auditors' recognition.

1. Establishing a Bachelor's program in auditing and/or establishing a Master or PhD program in auditing.
2. The auditing course credits should be classified in two categories of practical and theoretical, as most researches have indicated that more experience would provide auditors with more accurate judgment and more recognition and, therefore, these two will certainly lead to better result if accompanied.
3. It is suggested to make a revision on the issue of "materiality" when the university's auditing textbooks are compiled. In addition to auditing principles, the

standards of other countries and the instructions of the professional associations of auditing and accounting should be correctly translated into Farsi and be taught to students along with Iranian standards.

Auditors' Continuous Training

The research findings showed that auditors' different job classes affect their recognition so that the group of shareholders, managers and the group of supervisors and chief auditors have no similar recognition of materiality level. The group of shareholders and supervisors, in general, had similar recognition but no appropriate consensus observed in quantitative levels. Only the supervisors and the managers had similar recognition of materiality levels. Thus, it seems that participating the in-service training courses, professional conventions and conferences can provide a way to uniform the recognition of different classes, because various declarations and judgments and non-uniformity of auditors' operation concerning a single issue is not accepted by auditing reports users and, also, the excessive continuation of this may lead to the auditing profession's invalidity and the society's non- confidence to auditing reports.

Revision of Professional Instructions

As the research results showed that the external auditors had identified the qualitative factors better than quantitative factors, the Iranian Auditing Instruction was not able to well guide the auditors in identifying quantitative materiality levels. Despite to most international standards having considered the net profit as the basis of materiality computation, the Iranian Auditing Instructions have represented the amount of total assets and income or sales. In research conducted by Vahidi Elizei, Rahdarian (2008) where the auditors were asked to allocate a rank to each quantitative factor, this factor obtained the sixth rank. Therefore, the associations and organizations involved in setting standards and the rules of professional conduct regulations should design instructions conforming to international standards. Because, the existence of such guidelines would result in uniformity in audit declarations in similar conditions as well as in the resolution of problems caused by a change in auditors, uniformity of auditing work scope and, finally, improved efficiency and effectiveness of audit profession.

Make use of Students' Researches

The auditing authorities and standard-setters are suggested to use the results of students' researches to eliminate the flaw in standard #32 (audit materiality) and to help auditors' obtain more recognition through materiality and the relevant indicators.

Suggestions for Future Research

To obtain a higher degree of reliability from the present research results in another research, it is suggested to make a review of the firms' audit files to observe the extent of the auditors' recognition of the materiality levels in practice. Also, it is suggested to conduct research about the followings:

1. Assessment of the internal auditors' recognition of materiality levels
2. Review of other factors affecting the materiality and auditors' judgment
3. Comparison of standards in Iran with standards in other countries and finding the relevant flaws
4. Review of materiality judgments in the financial statements listed on the Stock Exchange
5. Examination of the effect of financial and non-financial information on the auditors' materiality judgment.

References

- American Institute Of Certified Public Accountants(AICPA). 1984.Audit Risk Materiality In Conducting On Audit .Statement On Auditingstandards No.47. New York, Ny:AICPA.
- Auditig standard ,technical committee of organization audit bulletin no 124
- Carpenter, B. W., and M. W. Dirsmith. 1992. Early debt extinguishment transactions and auditor materiality judgments: A bounded rationality perspective. *Accounting, Organizations and Society* 17 (8): 709-740.
- Chewning, G., K. Pany, and S. Wheeler. 1989. Auditor reporting decisions involving accounting principle changes: Some evidence on materiality thresholds. *Journal of Accounting Research* 27 (1): 78-96.
- Cpa journal,june2006. http://findarticles.com/p/articles/mi_qa5346/is_200606/ai_n21393212
- Estes, R., and D. D. Reames. 1988. Effects of personal characteristics on materiality decisions: A multivariate analysis. *Accounting & Business Research* 18 (72): 291-296.
- Financial Accounting Standards Board. 1980. *Qualitative Characteristics of Accounting Information*. Statement of Financial Accounting Concepts No. 2. Stamford, CT: FASB.
- Hsas yeganeh , yahya.kasiri,hossain 2003 "concept of materiality in audit and thair effect on auditors opinion .hesabras journal no 19
- International Accounting Standards Board (IASB). *International Accounting Standard (IAS) 1 - Presentation of Financial Statements*.
- Krogstad, J. L., R. T. Ettenson, and J. Shanteau. 1984. Context and experience in auditors' materiality judgments. *Auditing: A Journal of Practice and Theory* 4 (1): 54-74.
- Mckee,Thomas e.2000,"current materiality guidance for auditors" , Norwegian research council
- Messier, W. F., Jr. 1983. The effect of experience and firm type of materiality/disclosure judgments. *Journal of Accounting Research* 21 (2): 611-618.
- Messier W. F., Jr and Bennie N. M and Eilifsen A. 2005 . A Review And Integration Of Empirical Research On Materiality:Two Decades Later. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=786688
- Securities and Exchange Commission (SEC). 1999. *Materiality*. SEC Staff Accounting Bulletin No. 99. (August 12). Washington, D.C.: SEC.
- Patlillo,the concept of materiality in financial Reporting .(new york FERF,1976).chapter 3
- Tuttle, B., M. Collier, and R. D. Plumlee. 2002. The effect of misstatements on decisions of financial statement users: An experimental investigation auditor materiality thresholds. *Auditing: Journal of Practice and Theory* 21 (1): 11-27.
- Zarei, hossain 1996 "cocept of materiality in audit reporting process" M.A theseis, Tehran university.

Small Scale Supplier Satisfaction: An Explorative Finding from Indian Manufacturing Industry

Manoj Kumar Mohanty

Asst. Manager-Project, Larsen & Toubro Limited, Kansbahal,
Sundargarh, Orissa, India – 770034
[Email: mnm-kbl@kbl.ltindia.com](mailto:mnm-kbl@kbl.ltindia.com)

Prof (Dr) P Gahan

Head – Business Administration Department,
Sambalpur University, Burla, Orissa, india

Abstract

Long term relationship only exists if the stake holders of the relationship share an output call satisfaction. In industrial buyer supplier relationship it is mandatory that buying organization or buyer should ensure selling organization's or seller's satisfaction within the boundaries of the strategic walls. In an integrated supplier relationship management (SRM) cycle supplier satisfaction plays a vital role for the continual and smooth running of the cycle. The changing trend of purchasing and dramatically changing supply chain relationships forcing the buyers to concentrate and ensure the supplier satisfaction. The role of small scale suppliers is substantial in providing the competitive edge to the manufacturing organizations. So it is advisable to keep the satisfaction level of the SSI suppliers high in order to meet the desired operational flexibility. The satisfaction level of small scale suppliers only can be improved if the buying organization understands the problematic areas of the SSI units and formulate their purchasing policy accordingly. This paper is trying to explore the satisfaction level of SSI suppliers with large scale selected manufacturing Indian industries acting as buyers. The satisfaction attributes and index of the SSI suppliers towards buyers and the buying organization measured by survey and structured interviewed method. **Key words:** Supplier satisfaction, Buyer supplier relationship, Small Scale industries, problems of small scale industry, Manufacturing Industry, Supplier satisfaction attributes. **Type of Paper:** Research paper.

Suppliers are the value creators for the organization. Suppliers have emerged as a value-adding partner in industrial relationships (Johnston and Lawrence, 1988). These values can be derived effectively if the buying organization keeps a long term strategic relationship with suppliers. Long term and sustainable partnership is the essence of the current business trend. One of the prominent drivers of this relationship is supplier satisfaction. Supplier satisfaction is defined as a supplier's feeling of fairness with regard to buyer's incentives and supplier's contributions within an industrial buyer-seller relationship as relates to the supplier's need fulfilment (Essig & Amann, 2009). Satisfied suppliers can substantially contribute to the manufacturing flexibility, production life cycle compression, cost advantage and innovations to the original equipment manufacturer (OEM). If we will analyze our business pattern we will find that we are too much concerned about our customer and very little about the supplier. Very few companies do supplier satisfaction survey but we must remember supplier satisfaction can be linked to supply

chain effectiveness, customer satisfaction, buyer supplier relationship and more over a management tool for strategic positioning of the supply chain. As seen by manufacturers, suppliers have become true partners of value creation as opposed to mere deliverers of parts (Ulaga and Eggert, 2006). Buyer-supplier relationship must work to get the technical and financial advantages by concentrating mutual strength and weaknesses.

When it comes to find the small scale suppliers satisfaction many parameters of the supplier selection criteria and expectations of the buyers needs to be reviewed very closely. In other words the role of buyer plays a major role in SSI supplier satisfaction. Buyers must review their expectations with the capacity and capability of the small scale supplier. Buying organization must understand the existing circumstances prevailing in small scale industries with which relationship is in progress. This paper also highlights the actual problems facing by the small scale industries and how buyers can derive the satisfaction parameters of the suppliers as well as formulate purchasing

policy so that the relationship can be stronger and SSI supplier's satisfaction index can be increased.

This research paper tries to explore the satisfaction attributes for small scale industries from the perspective of large scale buying organization along with its measurement. The objective of this paper also how the management can use the survey as an management tool to improve the internal as well as external functions exposed to supply chain process. From the results the gap analysis can be made between suppliers and buying firm's mutual expectations. The quality relationship between buyer and supplier determines the strength of the supply chain. Basically the supplier satisfaction survey should a continuous process and it must continuously improve the supply chain and firm's competitive edge.

Literature Review

Satisfaction defined as "a positive affective state resulting from the appraisal of all aspects of a firm's working relationship with another company" in business relationship (Anderson and Narus, 1984). Satisfaction something a supplier's good feelings towards the original equipment manufacturer (OEM). This feeling created may be single time or multiple time interaction with buyer's or may be with the cross functional members of the organisation. As suppliers are emerging as value adding partners to the supply chain their satisfaction must be part in supply chain path. Supplier satisfaction which is something not discovered by many companies and it starts with individual's attitude towards suppliers (Donath, 1991). The manufacturer and its suppliers, vendors, and customers – that is, all links in the extended enterprise – working together to provide a common product and service to the marketplace that the customer desired and is willing to pay for throughout the life cycle of the product and service (Kuglin, 1998). So supplier selection plays a major role in placing the buyer's organisation in proper market place. Select suppliers who can best meet your needs from a cost and technology stand point and might mean the difference between choosing leading edge versus trailing-edge technology (Riggs & Robbins 1998).

It is economically reasonable to focus on supplier satisfaction, since supplier satisfaction indicates the quality of the buyer-seller relation-ship from the supplier's perspective (Essig, Amann, 2009). As a result research related to the study of establishing, developing and maintaining successful business relationships between customer and supplier firms has grown considerably.

Within this domain, a construct that has received particular attention within the domain of b-to-b relationships is trust (Ganesan, 1994). Trust, by infusing predictability (through enhanced information sharing) and mutual understanding into a relationship, provides a cushion against uncertainty (Luhmann, 1995).

When one organization try to understand the level of supplier satisfaction through any medium, the buying company basically sends a strong signal of trust, tries to improve and strengthen existing relationship (Spence, 1976). Trust and commitment are not only drivers of supplier satisfaction; they are stated to be the premises of relationship quality, which, in turn, can be classified as the object of a supplier's satisfaction statement (Walter, 2003; Maunu, 2003). Communication and feedback, Communication of requirements, Customer credibility, Early supplier involvement and purchasing power are the drivers of effective buyer supplier relationships (Lascelles and Dale ,1989). These are also factors of supplier satisfaction. Loyalty is a focal point in a long-term relationship, implying both a favourable attitude and customer relation. Satisfaction evolves as a consequence of one party's experience with the other's ability to fulfil norms and expectations (Harald Biong ,1993). When combining those two the relationship between supplier and retailer should be seen in a long-term perspective. The greater the satisfaction with the supplier, the more loyal the retailers are expected to be.

According to Biong's (1993) profitability and support have a strong impact on satisfaction and loyalty. Business value chain and the total cost of ownership linked together and the importance to each other, including the supplier value chain. Business value chain thinking has benefit both in money and time, which can be considered as elements of the supplier satisfaction concept also (Riggs & Robbins 1998). Supplier satisfaction measurement gives us opportunity to develop supplier relations in a way that both parties are willing and capable to do it. This approach is supported by Wong (2000). Wong's (2000) study explores the role of suppliers in improving customer satisfaction and finds that companies can make use of their suppliers in achieving high customer satisfaction. (Hult , 1997) aimed to reflect the two dimensions of supplier satisfaction. These are three cognitive (reputation, skill and transaction specific investments) and three behavioural (communications, coercive, power and flexibility)

Within the SME domain it is argued that most SMEs who want to build relationships often lack track record in terms of notable credibility signalled by reputation and character

endorsements (Prahalad and Hamel, 1990) According to Blois (1999), flexible conditions enable the creation of tolerance towards behavioural and environmental uncertainty and engender trust, which in turn gives room to the relationship to continue. There is positively co-relation between cooperation and satisfaction (Dwyer, 1980). The cooperation factor results in to trust, commitment and efficiency among buyer and supplier. Service quality and satisfaction linked with each other (Parasuraman, 1994). The behavioural attributes of manufacturing organisations also deliver satisfaction (Moorman, 1992). Firms must understand their chain partners in all respects, including comprehension of the sources, imbalances, and consequences of power such that the most beneficial use (or disuse) of this power can be directed to achieve supply chain performance and member satisfaction (Benton, Maloni, 2005)

Brief Back Ground of Indian Small Scale Industry

As per Indian small scale industry act, 2006 “ An industrial undertaking in which the investment in fixed assets in plant and machinery whether held on ownership terms on lease or on hire purchase does not exceed Rs 50 million is called a small scale industry.

Table.1– Definition of Micro, Small and Medium Industry

	Investment in plant and machinery/equipment (excluding land and building)	
	Manufacturing Enterprises	Service Enterprises
Micro	Up to Rs. 25 Lakh	Up to Rs. 10 Lakh
Small	More than Rs. 25 Lakh and up to Rs. 5 crore	More than Rs. 10 Lakh and up to Rs. 2 crore
Medium	More than Rs. 5 crore and up to Rs. 10 crore	More than Rs. 2 crore and up to Rs. 5 crore

Source: Annual Report (2006-2007)of Ministry of Small scale Industries, India

Small Scale Industries (SSI) in India has it's own substantial contribution to the socio-economic development of the country. SSI account for 95 per cent of the industrial units in the country. SSI make significant contributions to the country's GDP, manufacturing output, exports, and employment generation. MSMEs contribute 8 per cent of the country's GDP, 45 per cent of manufactured output, and 40 per cent of exports (Indian economic survey,2011).

SSIs existed in India for a long time in various sectors and contributed significantly in bringing down regional imbalance; generating employment opportunities, output, and exports; fostering entrepreneurship; in accelerating economic development. It occupies a position of prominence in India and contributes over 50% of the industrial production in terms of value-addition (Anand and Murugaiah, 2007). It has been estimated that a lakh rupees of investment in fixed assets in the small scale sector produces 4.62 lakhs worth of goods or services with an approximate value addition of ten percentage points and approximately employment for 4 persons (Indian economic survey,2011)

Table. 2 - Status of SSI as on 2005-2006

Units	12.3 Million
Employment	29.5 Million
Production	Rs. 4762.01 Billion
Industrial Production	40% of total production
Export	35%
GDP contribution	7%

Source: Annual reports of Ministry of Small scale industry , 2006-07

Table. 3 – Growth rate of SSI units

Year	Growth rates of 1970 base IIP	Growth rates of 2001-02 base IIP	Growth rates of manufacturing sector with base year 1993-94
2002-2003	7.68%	8.68%	6
2003-2004	8.59%	9.64%	7.4
2004-2005	9.96%	10.88%	9.2
2005-2006	10.40%	12.32%	9.1

Source: Annual reports of Ministry of Small scale industry, 2006-07

Table. 4– Growth rate of production

Year	Growth rates of SSI sector	Overall Industrial Sector
2002-2003	8.68%	5.70%
2003-2004	9.64%	6.90%
2004-2005	10.88%	8.40%
2005-2006	12.32%	8.10%

Source: Annual reports of Ministry of Small scale industry, 2006-07

The above figures clearly demonstrate the importance of small scale industries in Indian economy. The major opportunities in small scale industries are less capital requirement, extensive government support both by state government as well as central government, easy funding by national and state level financial institutes, liberal tax systems in material procurement, machinery procurement, Export promotions and training programs by government. This aspect helps in creating entrepreneurs, creates opportunity for engagement of Indian manpower, export opportunity and more over creates social balance.

Therefore it is more important for managers and the buyers of the buying organisations to look very carefully to this segment of business and if required the purchasing policy should be made in line with the country's SSI policy formulates by the Indian government. This effort will be easy if SSI supplier segment's policy, vision and mission can be made in line with the large scale manufacturing industry's policy, mission and vision. So it goes without mentioning that the satisfaction level of this segment of suppliers should be kept high. The desired course of action should be formulated accordingly.

Problems of Small Scale Industry (Research Zone)

The problems described here are derived from the small scale industries of western parts of Orissa, a fast industrially developing state of India. We have taken the views of the owners of the small scale suppliers (Manufacturing Units) to derive the basic problems faced by them. Surprisingly the problems faced by them are found to be almost same with the problems faced by other parts of the country and witnessed by previous researchers worked in SSI units.

Capital investment and inadequate working capital for operations is the major challenge of the SSI units. Lack of credit worthiness of this sector of industry does not allow them to take easy financial support from financial institutions. Very minimum number of SSI unit able to take cash credit facility from near banks. Delayed payment from the principal unit makes their position more worse. Even if there is business opportunity or demand of production they are not able to expand the manufacturing facility quickly. At times they borrow from the local money lenders to provide wage to their employees and exposed to exploitations.

Labour is the second most problematic area for SSI manufacturing units. The research zones, i.e. western parts of Orissa are falling under tribal area. Availability of skilled man power is seriously falling short in this zone.

The persons those who are technically qualified and having minimum skill level they don't prefer to work in small scale units. The reason being, the low remuneration and facility offered by SSI units. As cultivation is one of the source of income in this area, the division of manpower also observed here. The turnover ratio is very high with respect to other industry segment of India. This results in poor quality of product, low rate of production and instability.

Procurement of raw material is also an issue to SSI manufacturing units. As the requirements of raw materials of specific grade or composition is very less in comparison to large scale industry, the SSI unit not able to derive benefit from the discount or bargaining process. The quality of raw materials from the local markets is also not as per the requirements of the standards desired by OEM companies. They also face abnormal pricing and volume problem from the traders.

Poor quality is the bi-product of lack of skilled manpower, inadequate machineries, non-availability of instruments. This is not due to capital shortage only. The raw material, training of manpower and attitude of the suppliers are also the contributors of the poor quality.

Technology used in the SSI manufacturing units is almost out-dated. They try to get the maximum profit with the existing facility and technology they have. The demand of the OEM organisations are to use latest tools and machineries, use of high end welding technologies and impact of new process technologies for casting. This gap in the SSI unit results in low productivity and poor quality.

Other problems of SSI manufacturing units in the research area are having acute problem of electricity. Power cut at the tune of 5 to 6 hours in the summer offsets the production schedule and productivity. Proper documentation ability is also lacking which results in delay in document processing in payment and conflict with buyers. Due to non-availability of structured cost management technique, the cost of the process material and bidding know-how is far from standard practice of industry. There is no formal marketing process available to this sector of business. Personal contact, reference and local presence are the three major source of getting order for SSI manufacturing units. Application of information technology (IT) infrastructure and related applications are seriously lacking in this sector. New entrepreneurs are somehow using IT as a business tool but older generation SSI owners don't use IT as a business tool. SSI representative don't have awareness in managing conflict if arise due to any of the reason with the OEM. Passing the liquidate damage (LD) to supplier even though OEM is not having a clause from it's principal customer.

Linking the Problems of SSI to Satisfaction Attribute

The problem areas were identified with the primary data collected from the owners of the SSI manufacturing unit,

2nd line persons of the SSI unit, buyers and subject experts. They are listed down and corresponding expectations from the OEM also collected from the source. How these points are linked to the supplier satisfaction attributes explained in a tabular form as bellow.

Table. 5 – Problems of SSI, Their Expectations & Linkage to Satisfaction Attribute

Sl. No.	Problem Area	SSI supplier's expectation from Buying organisation (OEM) for solving the Problem	Linked to which attribute of supplier satisfaction
1	Capital Investment	(1)Release of payment in due time / before time. (2) Financial assistance as and when required with a scope for adjustment in pending bills. (3) provision for progressive payment	Order Management, Financial support
2	Raw material Procurement	(1)Supply of desired raw material for the finish product (2) Help them to get the volume discount as OEM by mediating with the traders/Supplying firm (3) Part payment of cost of raw material	Partnership Approach, Order management, Commitment
3	Labour	(1)Providing source of labour (2) Providing hands on training inside OEM (3)Extending some of the basic facility like group insurance (4)Training on safety and environment	Training, Mutual awareness, Communication
4	Technology	(1)Training on new technology adopted by OEM (2) Helping in implementing the new technologies with low cost (3) providing financial and managerial skills to adopt new technology (4) providing way to getting more with existing technology	Operational flexibility, Skill enhancement
5	Quality	(1)Communicate in structured way what exactly expected (2) Process / method guidelines for improvement in quality (3) Exposure to new measuring tools and instruments (4) Guidance from quality assurance department regularly	Communication, Training, Quality of service, Process enhancement
6	Electricity	(1)Financial assistance to have generators (2) Accommodate delay during extensive power cut periods	Mutual understanding, Financial support
7	Documentation	(1)Training on documentation (2)Providing Understanding what exactly required by OEM (2) minimize the documentation process	Cooperation, Training, Flexibility
8	Costing Structure	(1)Helping them understanding cost structure and how to calculate (2) In case of low bidding by supplier advise them to rectify	Purchasing ethics, order management, attitude, loyalty, trust
9	Training	(1)Training required without any cost in every step where OEM is trying to implement process, product etc.	Partnership approach, trust
10	Information Technology	(1)Provision of provide IT infrastructure or training in supply chain applications (2) Establishing link work stations between supplier end and OEM	Cooperation, system enhancement, process improvement
11	Conflict	(1)Avoidance of conflict situation (2) Grievance handling cell for proper justice	Mutual understanding, faith, trust
12	Liquidate Damage (LD)	(1)Allowing of sufficient /actual production time , so that LD should not be applicable to them.	Order Management

Source: Author; by primary survey, 2011

Findings from Literature & Bridging the Gap

Based on literature reviews and the previous research works carried out by researchers we can sum-up the following attributes of supplier satisfaction. They are trust, commitment, understanding, innovation, flexibility, communication, reputation, coercive and non-coercive power, cooperation, bonds, dependency, quality of service, mutual awareness, and behaviour of buying organisation, loyalty and attitude.

Still many of the core issues attached with the small scale supplier's satisfaction not addressed in the previous researchers. These gaps were found with the interaction with the SSI suppliers those who are attached to large scale manufacturers for a longer time, structured discussions with buyers those who are having 10+ years of experience in handling small scale suppliers and opinion from industry experts. The factors derived from the primary survey and not covered by previous researchers on SSI supplier satisfaction are Partnership approach, Process / product improvement at supplier's end, skill enhancement of supplier's manpower, capital support, providing safety and environmental know-how, helping in raw material procurement, enabling them for proper documentation, technical and behavioural training, helping in managing technological up gradation and implementation, guidance for cost structure, advise for proper bidding value and avoidance of conflict situation. Liquidate damage clause management is also needs proper attention from OEM buyers. Another two factors, i.e. performance feedback and policy change feedback to suppliers from OEM plays role in their satisfaction also needs to be tested.

Based on the previous research, primary survey on the SSI suppliers problem and their expectations, experts from the SCM area, with the help of experienced buyers the questionnaire prepared (attached as Annexure-I) to test the satisfaction index of the SSI manufacturing units with respect to the selected Indian manufacturing units.

Research Methodology

This research work carried out in the manufacturing industries of Orissa, India. To avoid the biasness we have taken survey of the existing small scale suppliers associated with four large scale Indian manufacturing industries. Sample selected based on the interaction of the persons from the supplier end to the buying organization. Many occasions the owners of the selling organization found to

be as sample and many occasions both the owner and the 2nd man of the selling organization found to be as sample. Initially 360 structured questionnaires sent for collection of data. Only a sample of size 245 is collected by E-mail, postal system and personal interactions depending upon the location of the supplier. A usable sample of 213 is considered for analysis, rest 32 are rejected for incomplete information. The structured questionnaire is developed based on the literature review, expert opinions and working experience of the authors. During the exploration of SSI unit's problems the experienced (10+ years) buyers views were also considered apart from the SSI manufacturing suppliers. The questionnaire is designed using five point Likert scale where 1 meant strong disagreement and 5 mean strong agreement with a particular attribute of supplier satisfaction. The scale reliability is found to be 0.764 which is found to be above the accepted level (Nunally, 1978). The Statistical Package for Social Sciences (SPSS version 18.0) was used to run the data collected from the supplier's end.

Results & Discussions

Table 6 contents the descriptive statistics of the opinion of 213 small scales manufacturing suppliers those who are exposed to supply chain practices and attached to large scale manufacturing industries through whom the supplier satisfaction survey conducted.

Table. 6 – Descriptive statistics of small scale supplier satisfaction attributes

Var	Variable Description	Mean	Std. Deviation	Analysi N
1	Adherence to payment terms	3.99	1.048	213
2	Process Improvement initiatives	3.25	1.036	213
3	Amount of Contribution to the Business Function	3.66	0.950	213
4	Early supplier involvement in design stage	3.98	0.566	213
5	Involvement in decision making process	2.70	0.735	213
6	Payment habits	3.53	0.509	213
7	Clarity of terms and conditions	3.64	0.481	213
8	Transparency in Purchasing procedure	2.84	0.626	213
9	Structured & Practical bargaining Procedure	3.24	0.928	213

Var	Variable Description	Mean	Std. Deviation	Analysi N
10	Guidance during First time order processing	4.15	0.565	213
11	Allowing of Delivery Period	3.01	0.755	213
12	Product development initiatives	2.68	0.741	213
13	Assurance of orders	3.19	0.902	213
14	Commercial clarity in order	3.26	0.820	213
15	Guidance from Quality control	2.79	0.793	213
16	Availability of conflict management cell	3.21	0.821	213
17	Technical clarity in order	3.02	1.02	213
18	Skill building attitude	2.58	0.732	213
19	Communication medium	3.14	0.713	213
20	Quality of interaction with buyer	3.34	0.811	213
21	Documented procedure for Conflict management	2.91	0.893	213
22	Capital Investment	4.22	0.478	213
23	Accommodative terms (If any)	3.28	0.754	213
24	Attitude of Cross functional team members	4.35	0.534	213
25	Frequency of communication	3.76	0.675	213
26	Range of Contracts	2.63	0.530	213
27	Benefit forwarding approach	3.58	0.707	213
28	LD forwarding Clause	3.72	0.886	213
29	Inspection Clauses	3.12	0.729	213
30	Guidance to avoid conflicting situation in future	3.63	0.520	213
31	Quality of Communication	2.93	0.841	213
32	Number of supplier meets	2.72	0.668	213
33	Attitude for solving	3.23	0.836	213
34	Degree of response	3.08	0.271	213

Var	Variable Description	Mean	Std. Deviation	Analysi N
35	Interaction with stores and billing section	2.62	0.759	213
36	Performance Feedback	3.09	1.041	213
37	Quality of Reaction	3.96	0.201	213
38	Policy Change Feedback	3.64	0.480	213
39	Response Time	3.38	0.852	213

To measure and reassure the internal consistency, reliability test was conducted on the 39 variables bearing a Cronbach's alpha of 0.764. The results inferred the questionnaire was measuring the small scale supplier satisfaction attributes in manufacturing industry in a meaningful way. Then factor analysis was used to remove the redundant/highly correlated variables from the survey data and to reduce the number of variables into definite number of dimensions. The factor analysis is performed using the principal component extraction method with varimax rotation. The eigen values of selected factors were greater than 1. Table 8 shows the total variance explained by the factors extracted.

Table. 7 - Cronbach's Alpha

Case Processing Summary			
		N	%
Cases	Valid	213	100
	Excluded	0	0
	Total	213	100
Reliability Statistics			
Cronbach's Alpha		N of Items	
0.764		39	

Listwise deletion based on all variables in the procedure.

Table. 8 - Total Variance Explained

Component	Initial Eigen values			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11.76	30.162	30.162	11.763	30.162	30.162	8.427	21.609	21.609
2	8.881	22.772	52.935	8.881	22.772	52.935	7.726	19.81	41.419
3	3.96	10.153	63.088	3.96	10.153	63.088	6.45	16.537	57.956
4	3.39	8.692	71.78	3.39	8.692	71.78	3.826	9.811	67.767
5	2.236	5.733	77.513	2.236	5.733	77.513	3.801	9.745	77.513
6	1.968	5.046	82.559						
7	1.266	3.247	85.806						

Component	Initial Eigen values			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
8	1.057	2.709	88.515						
9	0.793	2.035	90.55						
10	0.599	1.535	92.084						
11	0.444	1.139	93.224						
12	0.372	0.953	94.177						
13	0.331	0.848	95.025						
14	0.266	0.682	95.707						
15	0.244	0.624	96.331						
16	0.194	0.497	96.828						
17	0.19	0.486	97.314						
18	0.153	0.391	97.706						
19	0.122	0.312	98.018						
20	0.116	0.297	98.314						
21	0.11	0.281	98.595						
22	0.076	0.194	98.79						
23	0.073	0.186	98.976						
24	0.064	0.163	99.139						
25	0.062	0.159	99.298						
26	0.054	0.139	99.437						
27	0.052	0.133	99.57						
28	0.041	0.105	99.675						
29	0.027	0.07	99.745						
30	0.024	0.062	99.808						
31	0.02	0.051	99.859						
32	0.014	0.035	99.894						
33	0.012	0.031	99.925						
34	0.009	0.023	99.948						
35	0.007	0.018	99.967						
36	0.005	0.012	99.979						
37	0.004	0.01	99.989						
38	0.003	0.008	99.998						
39	0.001	0.002	100						

Extraction Method: Principal Component Analysis

In the initial application, the numbers of variables were reduced from 39 to 38. In the second application, these 38 variables were classified under five dimensions based on their factor-loading score. The sorted rotated values of factor loading with minimum value of 0.5 or more are considered. After dropping the redundant variables having all the five components loading less than 0.5 the following matrix has been formed to understand the significant

components that explain 77.51% of variance. Generally, factor loading represents how much a factor explains a variable. The variables which was dropped is quality of reaction in conflicting situation. High loading indicates that the factor strongly influences the variable. Assuming a factor loading of more than 0.80 as having high impact on the variables, it is concluded that some variables which are less than 0.80 need more attention.

Table. 9 - Rotated Component Matrix

SI	Variables	Component				
		1	2	3	4	5
1	Adherence to payment terms	0.841	-0.09	0.176	0.24	0.014
2	Process Improvement initiatives	0.408	0.507	0.237	-0.267	0.272
3	Amount of Contribution to the Business Function	-0.451	-0.195	-0.013	0.565	0.442
4	Early supplier involvement in design stage	-0.402	0.294	-0.276	0.627	-0.02
5	Involvement in decision making process	-0.105	0.416	-0.042	0.613	-0.065
6	Payment habits	0.708	-0.356	-0.382	0.23	-0.205
7	Clarity of terms and conditions	-0.553	-0.447	0.176	-0.035	-0.225
8	Transparency in Purchasing procedure	0.632	0.303	-0.224	-0.33	0.083
9	Structured & Practical bargaining Procedure	0.86	-0.057	-0.138	0.375	0.179
10	Guidance during First time order processing	0.177	-0.916	-0.061	0.187	-0.156
11	Allowing of Delivery Period	0.742	-0.498	0.272	0.105	0.046
12	Product development initiatives	0.2	0.766	0.347	0.244	-0.113
13	Assurance of orders	0.934	-0.055	-0.013	-0.014	0.144
14	Commercial clarity in order	0.552	-0.238	-0.433	0.211	0.227
15	Guidance from Quality control	0.401	-0.552	0.028	0.122	0.093
16	Availability of conflict management cell	-0.438	0.381	-0.35	0.16	0.522
17	Technical clarity in order	0.736	0.439	-0.118	0.119	0.343
18	Skill building attitude	0.287	0.64	-0.11	0.341	-0.021
19	Communication medium	0.482	-0.463	0.508	0.09	-0.074
20	Quality of interaction with buyer	0.856	0.022	0.01	0.142	0.154
21	Documented procedure for Conflict management	0.126	0.53	0.044	-0.413	0.634
22	Capital Investment	0.35	-0.794	0.154	0.245	0.254
23	Accommodative terms (If any)	-0.828	-0.197	0.395	-0.056	-0.035
24	Attitude of Cross functional team members	0.466	-0.539	0.391	-0.109	-0.198
25	Frequency of communication	-0.491	-0.41	0.676	-0.178	-0.068
26	Range of Contracts	0.063	0.065	0.144	0.818	-0.215
27	Benefit forwarding approach	-0.558	-0.163	0.473	0.019	0.366
28	LD forwarding Clause	-0.663	0.461	0.43	0.183	0.262
29	Inspection Clauses	0.574	0.163	-0.308	-0.202	0.078
30	Guidance to avoid conflicting situation in future	-0.495	-0.379	0.145	0.368	0.59
31	Quality of Communication	0.369	-0.482	0.586	0.322	0.128
32	Number of supplier meets	0.275	0.562	0.397	0.34	-0.144
33	Attitude for solving	-0.282	0.465	0.328	0.148	0.699
34	Degree of response	-0.295	0.208	0.254	0.704	-0.311
35	Interaction with stores and billing section	0.418	0.049	0.745	-0.304	0.289
36	Performance Feedback	0.432	0.823	0.227	-0.123	-0.014
37	Quality of Reaction	-0.212	0.106	0.062	-0.023	0.145
38	Policy Change Feedback	-0.382	-0.608	-0.221	0.332	0.157
39	Response Time	0.188	0.374	0.542	0.268	-0.026

Extraction Method: Principal Component Analysis

Based on the results of factor analysis, the variables were classified into five dimensions, which were suitably named. The dimensions and the corresponding variables are shown table 10. The analysis revealed the following dimensions:

Table. 10- Small Scale Manufacturing Supplier satisfaction factors

SL	FACTORS	SUB-FACTORS	
1	ORDER MANAGEMENT	1.1	Transparency in Purchasing procedure
		1.2	Clarity of terms and conditions
		1.3	Adherence to payment terms
		1.4	Payment habits
		1.5	Allowing of Delivery Period
		1.6	Structured & Practical bargaining Procedure
		1.7	Inspection Clauses
		1.8	Technical clarity in order
		1.9	Commercial clarity in order
		1.10	Quality of interaction with buyer
		1.11	Assurance of orders
		1.12	Benefit forwarding approach
		1.13	Accommodative terms (If any)
		1.14	LD forwarding Clause
2	PARTNERSHIP APPROACH	2.1	Process Improvement initiatives
		2.2	Product development initiatives
		2.3	Guidance from Quality control
		2.4	Skill building attitude
		2.5	Performance Feedback
		2.6	Policy Change Feedback
		2.7	Number of supplier meets
		2.8	Guidance during First time order processing
		2.9	Attitude of Cross functional team members
		2.10	Capital Investment
3	COMMUNICATION	3.1	Communication medium
		3.2	Frequency of communication
		3.3	Quality of Communication
		3.4	Response Time
		3.5	Interaction with stores and billing section
4	STRATEGIC IMPORTANCE	4.1	Degree of response
		4.2	Amount of Contribution to the Business Function
		4.3	Early supplier involvement in design stage
		4.4	Involvement in decision making process
		4.5	Range of Contracts

SL	FACTORS	SUB-FACTORS	
5	CONFLICT MANAGEMENT	5.1	Availability of conflict management cell
		5.2	Documented procedure for Conflict management
		5.3	Attitude for solving
		5.4	Guidance to avoid conflicting situation in future

Factor 1: Order Management

Factor 1 stands for order management. Proper and effective order management is the basic rule of obtaining supplier satisfaction. This statement found to be very correct with this supplier satisfaction survey. Out of 39 variables, 14 variables fall under the order management practices by buying organisations. The most influencing attributes found as Assurance of orders with a factor loading of 0.934, Adherence to payment terms with a factor loading of .841, Structured bargaining process with factor loading 0.860, Quality of interaction with buyer factor loading of 0.856 and Accommodative terms and condition with factor loading of 0.828. Other attributes under this factor are Transparency in Purchasing procedure, Clarity of terms and conditions, Payment habits, Allowing of Delivery Period, Inspection Clauses, Technical clarity in order, Commercial clarity in order, Benefit forwarding approach, LD forwarding Clause. One thing very clearly emerging from the study that the payment related issues and fair practices in the purchasing gives most satisfaction to the small scale manufacturing suppliers. The interaction with the buyer with a qualitative way also creates satisfied supplier as well as makes buyer supplier relation stronger.

Factor 2: Partnership Approach

Factor 2 stands for partnership approach of the buying organisation towards small scale suppliers. Current supply chain practices, giving more emphasis on integration and partnership. This survey also gives indication that through partnership approach supplier satisfaction can be achieved. Out of 39 variables, 10 variables are falling under this factor. The high influencing factors are Guidance during First time order processing with factor loading of 0.916, Capital Investment with factor loading of .794, Performance Feedback with factor loading .823 and Product development initiatives with factor loading .766. Other attributes are Process Improvement initiatives, Guidance from Quality control, Skill building attitude, Policy Change Feedback, Number of supplier meets and Attitude of Cross

functional team members. It appears from the results that during the first time order processing the guidance by the concerned buyer creates a major impression on the supplier. Performance feedback to the supplier helps them to improve and they take positively to this factor. Product development initiates at supplier end creates high motivation and feel assure of business growth and order assurance.

Factor 3: Communication

Factor 3 represents communication between the buying organisation and the supplier's organisation. Usually communication takes place between buyer and seller. The attributes falling under this factor are Communication medium, Frequency of communication, Quality of Communication, Interaction with stores and billing section, Response Time with a factor loading of 0.508, 0.676, 0.586, 0.745 and 0.542. This clearly indicates the suppliers contacting points except buyer i.e. stores and billing section plays a role in supplier satisfaction with a factor loading of 0.745. The interaction of the concerned persons in stores and billing section, their attitude towards supplier plays role in building relationship as well as satisfaction of small scale suppliers.

Factor 4: Strategic Importance

Factor 4 represents strategic importance, what are the strategic factors involved in the supplier satisfaction aspect and how suppliers are having their opinion on it. The high influencing attributes are Range of Contracts with factor loading of 0.818 and Degree of response with factor loading of 0.704. Other attributes are Amount of Contribution to the Business Function, Early supplier involvement in design stage, Involvement in decision making process with a factor loading of 0.565, 0.627, 0.613 respectively. This clearly indicates that the more the involvement in business areas more is the satisfaction. Suppliers feel like a associate business partner by so. Early supplier involvement and involvement in decision making processes are having their own impact on supplier satisfaction.

Factor 5: Conflict Management

Factor 5 stands for conflict management. During business process where the mutual expectation does not meet in supply chain conflicting situation arises between buyer and supplier. But there must some structured way of solving the situation so that relationship does not hamper and

satisfaction of the suppliers retains. The attributes surfaced in this factor are Availability of conflict management cell, Documented procedure for Conflict management, Attitude for solving, Guidance to avoid conflicting situation in future with a factor loading of 0.522, 0.634, 0.699 & 0.590 respectively. In very rare industries the documented procedure and structured way of handling conflict found.

Conclusion and Recommendations

There is no end to expectations; still in our business relationship we can satisfy the upstream and the downstream with the limited resource and within the strategic boundaries of our organization. Today's business and the supply chain effectiveness largely depend upon the integration and satisfaction of the suppliers associated with the business. From this perspective the supplier satisfaction survey has its own importance. This helps the buying organization or buyers as a tool to guide the policy and purchasing practices. This also helps to minimize the dissatisfaction among suppliers and new direction to retain the existing suppliers. Satisfaction and dissatisfaction of suppliers also linked to the firm and buyer's performance. The satisfied supplier usually produce high quality product and provides high delivery index. This will lead to higher profitability of the buying organization. The satisfaction also leads to stronger buyer supplier relationship with ultimately results in firms flexibility and operational proficiency.

From the survey results it is quite clear that order management process and partnership approach attributes are the major factors which create the satisfaction among small scale manufacturing suppliers. out of the factors order assurance, payment term adherence, quality interaction of from the buyer, first time guidance to the order processing, range of contracts, performance feedback and product development are the most influencing factors. so the buying manufacturing industry must look closely to this factors and formulate action policies so that the satisfaction level of the small scale suppliers retained and improved. In the problem exploration section also we have found that technology, labor, capital, raw material procurement are the core problem areas for the small scale industry. if large scale organizations can influence those factors in the favor of small scale suppliers, loyalty and long term business relationship can be created which is going to give the competitive edge to the business. the other aspect is also attached to this i.e. local supplier development, locality development and social balance. one important thing the survey conducting organization should remember that the

fact should come out, by not exposing the suppliers. If possible the survey must be conducted by any third party agency so that clarity and integrity of the survey remains.

Managerial Implications

Supplier satisfaction survey has many angles of utility from the buying organizations perspective. Firstly it indicates the present state of affairs. The organizational policy, purchasing practice, buyer supplier relationship, vendor development initiatives practically exposed with the correct survey. One thing must be ensured during the survey that, the suppliers, should be allowed to disclose exact dissatisfaction without disclosing their identity. Secondly potential areas of improvement can be easily identified with the survey. Proper action plan can be made to improve the internal operational system as well as external operational system. Thirdly it gives direction to improve the buyers and suppliers relationship. It closes the loop holes of the dissatisfaction of suppliers. It also clears the perceived or actual conflicts between buyer and supplier.

Supplier satisfaction survey has been ignored by many manufacturing industries till date. Those companies who are doing supplier satisfaction survey, they do it by themselves, which may not clearly indicate the actual results. The survey by third party agencies may reveal the facts. Supplier survey also provides stress to the purchasing department. But this is only for the purchasing process improvement and system alignment. Little biasness may occur due to specific conflicting cases, but by and large supplier satisfaction survey is a must to the organization those who are involved in the supply chain processes. Supplier satisfaction survey must be carried out one in a year. For specific conflicting cases individual supplier's feedback also can be taken for further clarity in the purchasing policy or system.

Scope for Future Research

Supplier satisfaction surveys just a beginning. No doubt it focuses on many issues like supplier productivity, manufacturing flexibility, buyer supplier relationship and loyalty. But many more can be explored with the help of such surveys. Supplier satisfaction across the different levels of supplier class i.e. Micro, Mini, Small and Medium industries can be studied. Supplier satisfaction index and the productivity levels also could be one of the research areas. Supplier satisfaction how linked to supplier loyalty can also be another area of research. The impact of supplier satisfaction survey on the purchasing policy and procedure

of the large scale industries also can be studied, so that standards can be formed across the industry. This study and research method may vary from industry to industry but general guidelines can be formed and organizations can be more and more benefitted by using supplier survey as a management tool.

References

- A Parasuraman, Zeithaml, V. and Berry L (1994), “Reassessment of expectations as a comparison standard in measuring consumer perceptions of service quality : Implications for future research”, *Journal of Marketing*, Vol, 58, January, PP. 11-24.
- Anand S, Murugaiah V, “small scale industries in india: an evaluation of performance in the post-liberalized scenario”, the iup journal of managerial economics, 2006, vol. iv, issue 4, pages 63-77
- Anderson JC & Narus JA (1984) A model of the distributor’s perspective of distributor-manufacturer working partnerships. *J Mark* 1984;48 pp. 62–74.
- Benton W. C., M. Maloni / *Journal of Operations Management* 23 (2005) 1–22
- Biong H (1993) Satisfaction and loyalty to suppliers within the grocery trade. *European journal of marketing* vol 27. no.7. pp. 21–38.
- Blois, K.J. (1999), “Trust in business-to-business relationships: an evaluation of its status”, *Journal of Management Studies*, Vol. 36 No. 2, pp. 197-215.
- Donath R (1991) Have you hugged supplier today? *Marketing News*. No.11. pp 4.
- Dwyer, F. R, “Channel member satisfaction: laboratory insights”, *Journal of Retailing*, Vol.56, Summer (1980), PP. 59-74
- Essig M, Amann M- *Journal of Purchasing & Supply Management* 15 (2009) P.103–113
- Ganesan, S. (1994), “Determinants of long-term orientation in buyer-seller relationships”, *Journal of Marketing*, Vol. 58 No. 2, pp. 1-19.
- Johnston, R., Lawrence, P., 1988. Beyond vertical integration. The rise of the value added partnership. *Harvard Business Review* 66 (4), 94–101.
- Hult, G.T.M., Neese, W.T. and Bashaw, R.E. (1997), “Faculty perceptions of marketing journals”, *Journal of Marketing Education*, Vol. 19 No. 1, pp. 37-52.
- Kuglin FA (1998) Customer-centered supply chain management: a link-by-link guide. Amacom. P.278
- Luhmann, N. (1995), *Social Systems*, Stanford University Press, Stanford, CA
- Lascelles D & Dales B (1989) The Buyer-Seller relationship on total quality management. *Journal of Materials Management*. Summer. pp. 10–19.
- Moorman, C., Zaltman, G. and Deshpande, R. (1992), “Relationships between providers and users of marketing research: the dynamics of trust within and between organisations”, *Journal of Marketing Research*, Vol. 29, August, pp.314-329.
- Riggs D & Robbins S (1998) The executive's guide to supply management strategies : building supply chain thinking into business processes. Amacom. 245 p.
- Spence, M., 1976. Informational aspects of market structure: an introduction. *Quarterly Journal of Economics* 90 (4), 591–597.
- Uлага, W., Eggert, A., 2006. Value-based differentiation in business relationships: gaining and sustaining key supplier status. *Journal of Marketing* 70 (1), P. 119–136.
- Walter, A.Muller, T.A., Helfert, G., Ritter, T., 2003. Functions of industrial supplier relationships and their impact on relationship quality. *Industrial Marketing Management* 32(2), 159–169.
- Wong A (2000), Integrating supplier satisfaction with customer satisfaction. *Total Quality Management*. Jul 2000. vol146. pp. S427–S432.

Section DET

Engineering and Technology

Section coordinated by the Department of Engineering and Technology in the School of Doctoral Studies of the European Union

Head of the Department:	<i>Professor Luidolf Bosch</i>
Deputy Head of Engineer:	<i>Professor Robert Munier</i>
Deputy Head of Technology:	<i>Dr. Professor József Ruppel</i>

Reviewers Body Coordinators:

Chair of Aerospace and Transport Engineering:	<i>Professor Bence Anzanberger</i>
Chair of Chemical Engineering:	<i>Professor Louis Maté</i>
Chair of Civil Engineering:	<i>Professor Kurt Nigge</i>
Chair of Electrical Engineering:	<i>Professor Michael Kirkbridge</i>
Chair of Mechanical Engineering :	<i>Professor Alex Nickols</i>
Chair of Computer and Software Engineering:	<i>Professor Thomas Aaronson</i>
Chair of Nanotechnology:	<i>Professor Jeffrey Dessler</i>
Chair of Mechatronics:	<i>Professor Sarah Poliza</i>
Agricultural Engineering and Food Technology:	<i>Professor Michael Leen</i>
Chair of Energy Technology and Engineering:	<i>Professor George Szentpetéri</i>

Engineering and Technology Section Content

Article 1	Page 78
Globalization and Energy Demands in the 21st Century <i>Author: Andrew R. Thorp (MSc)</i> <i>Country: Germany</i>	
Article 2	Page 85
Overview and Analysis on Cyber Terrorism <i>Author: Steve Saint-Claire (MSc)</i> <i>Country: Australia</i>	
Article 3	Page 99
The impact of computer aided design on human factors in aviation safety <i>Author: Anne Curtney (MSc)</i> <i>Country: USA</i>	

Globalization and Energy Demands in the 21st Century

Andrew R. Thorp (MSc)

Master of Science and candidate to PhD in Engineering at the Department of
Engineering and Technology of the School of Doctoral Studies of the EU

[Email: ar.thorp@det.sds.eu](mailto:ar.thorp@det.sds.eu)

Abstract

According to a May 2004 report from Deloitte Research, supplying enough energy on a reliable basis at prices that will not cripple the global economic growth has become a challenge with consequences that are difficult to predict (Globalization pp). Although this will provide new opportunities for oil and gas companies, pipelines, generators, utilities and others in the energy business, it also carries serious risks (Globalization pp). The demand for energy is growing, not only in the developed economies of Europe, Japan and North American, but in developing countries as well (Globalization pp). In fact, the fastest demand growth is in China and other emerging markets, thus from one side of the globe to the other, societies are needing and demanding more fuel (Globalization pp). **Key words:** Energy, Globalization, Energy Demand.

Twenty years ago, globalization was barely discussed, because at the time, less than fifteen percent of the world's population participated in true global trade (Marber pp). Third World countries were mere pawns in the Cold War's global chess game, and the prospect of the Soviet Union or Communist China actually integrating with the West economically, or regimes in Latin America or Asia abandoning central planning, seemed remote and improbable (Marber pp). In fact, the possibility of any of these countries making any meaningful socioeconomic progress and reaching Western standards of living were considered completely unrealistic (Marber pp).

On average, people are living twice as long as they did a hundred years ago, and the world's aggregate material infrastructure and productive capabilities are hundreds, if not thousands, of times greater than they were a century ago (Marber pp). Moreover, much of this has occurred since the last half of the twentieth century, with a powerful upsurge during the last twenty-five years (Marber pp). In the last two generations, there have been gains in virtually every meaningful aspect of life and the trend will likely continue upward at least through the middle of the twenty-first century (Marber pp). The fact that people are living longer, fuller lives is most evident in the Third World countries, where during the last fifty years, the life expectancy has increased by more than fifty percent, reaching levels the West enjoyed only two generations ago (Marber pp).

The world today has more educated people with greater intellectual capacity that at any time in history, and this is

particularly evident in much of Asia, where mass public education has enabled billions of people to increase their productivity and integrate as workers and consumers into the global economy (Marber pp). These same trends can be found in Eastern Europe and in part of Latin America as well, leading to historic highs in economic output and financial assets per capita (Marber pp). During the twentieth century, economic output in the United States and other West European countries often doubled in less than thirty years, and Japan's postwar economy doubled in less than sixteen years (Marber pp). Just in the last few decades, the economies of developing countries have grown so quickly that some, like South Korea during the 1960's and 1970's, and more recently China, have often doubled productive output in just seven to ten years (Marber pp).

It must be remembered that poverty was the human living standard for most of recorded history, and until about two hundred years ago, virtually everyone lived at a subsistence level (Marber pp). In 1931, economist John Maynard Keynes wrote in "Essays in Persuasion," that from the earliest times of recorded history to the beginning of the eighteenth century, "there was no very great change in the standard life of the average man living in civilized centers of the earth" (Marber pp). Although there were golden intervals, there was no "progressive violent change," due to two reasons, the absence of technical improvements and the failure of capital to accumulate (Marber pp). However, beginning in the early nineteenth century, the proportion of the world's population living in poverty declined from over

eighty percent in 1820 to under fifteen percent in 2000, even as the world's population exploded from over one billion to more than six billion (Marber pp).

The root of modern prosperity can be found in the application of mass production technology, together with excess capital and a free market to exploit such technologies (Marber pp). The shifting U.S. labor pattern from low-wage agricultural labor to manufacturing to higher-paid office and service employment during the last two centuries resulted largely from trade, and similar shifts are now seen all over the globe (Marber pp). For example, during the 1950's and 1960's, the U.S. imported electronics from Japan and exported cars, then in the 1970's, the U.S. began importing small cars from Japan, and the last thirty-odd years has seen Japan lose its dominance in electronics and economy cars due to competition from China and South Korea, resulting in Japan shifting to more expensive luxury cars and sport utility vehicles (Marber pp). Although jobs were lost, gained and relocated in the U.S. and abroad during these market shifts, the living standards in United States, Japan, South Korea, and China have all improved dramatically over the same time span (Marber pp). According to the United Nations, two-thirds of the world's middle-class citizens in 1960 lived in the industrialized world, the United States, Canada, Western Europe, Japan and Australia, then by 1980, over sixty percent of the global middle class lived in developing countries, and by 2000, the number had reached eighty-three percent (Marber pp). It is predicted that India and China combined could easily produce middle classes of 400-800 million people over the next two generations, roughly the size of the current middle-class populations of the United States, Western Europe and Japan combined (Marber pp).

There is no avoiding the fact that the success of globalization is underscored by dramatic increases in consumption, and with increased consumption comes environmental degradation (Marber pp). Current and projected damage to the environment can impede economic progress, and climatic changes attributed to greenhouse gas emissions and pressure on natural resources have become serious problems (Marber pp). Resource scarcity is an issue the world will have to confront as two to three billion more people consume like middle class Americans over the next fifty years (Marber pp). Globalization of the energy market is deepening and broadening through international trade as well as through "cross-investments, deregulation of domestic markets, and industrial restructuring that links the older energy industries to the new global political

economy" (Harris pp). This transformation of energy industries and markets can be seen around the world and offers great potential in terms of economic efficiency, technology development and consumer choice (Harris pp).

According to a report published in March 2005 by the National Intelligence Council, growing demands for energy, especially by the rising powers, over the next fifteen years will have substantial impacts on geopolitical relations (US pp). The single most important fact affecting the demand for more energy will be global economic growth, particularly that of China and India (US pp).

Although the trend is toward more efficient energy use, the report predicts that total energy consumed will likely rise by fifty percent over the next two decades compared to a thirty-four percent expansion during the last twenty years of the twentieth century, with an increasing share provided by petroleum (US pp).

Renewable energy sources such as hydrogen, solar, and wind energy will probably account for only about eight percent of the energy supply in 2020 (US pp). Russia, China, and India all plan expansions of their nuclear power sector, however, nuclear power will likely decline globally in absolute terms over the next decade (US pp).

The International Energy Agency predicts that with "substantial investment in new capacity, overall energy supplies will be sufficient to meet growing global demand" (US pp). However, continued limited access of the international oil companies to major fields may restrain this investment, and many of the areas, such as the Caspian Sea, Venezuela, West Africa and South China Sea, that are being counted on to provide increased output involve substantial political and economic risk (US pp). Moreover, traditional supplies in the Middle East are also increasingly unstable, thus sharper demand driven competition for resources, accompanied by major disruption of oil supplies, is among the key uncertainties (US pp).

China and India lack adequate domestic energy resources and will have to ensure continued access to outside suppliers, thus the need for energy will become a major factor in shaping the foreign and defense policies of these countries, including expanding naval power (US pp). According to experts, China will need to boost its energy consumption by roughly 150 percent and India will need to almost double its consumption by 2020 in order to maintain a steady rate of economic growth (US pp). Beijing's growing energy needs are likely to force China to increase its activist role in the world, especially in the Middle East, Africa, Latin America, and Eurasia (US pp). Yet, in trying

to maximize and diversify its energy supplies, China fears being vulnerable to pressure from the United States which Chinese officials view as having an aggressive energy policy that can be used against Beijing (US pp). For more than ten years, Chinese officials have stated that production from Chinese firms investing overseas is more secure than imports purchased on the international market (US pp). In order to secure more reliable access, Chinese firms are being directed to invest in projects in the Caspian region, Russia, the Middle East and South America (US pp).

The National Intelligence Council's report also states that Europe's energy needs will probably not grow to the same extent as those of the developing world, partly because of Europe's expected lower economic growth and more efficient use of energy (US pp). Europe's increasing preference for natural gas, combined with depleting reserves in the North Sea, will provide an added boost to political efforts that are already under way to strengthen ties with Russia and North Africa, since gas requires a higher level of political commitment by both sides in designing and constructing the necessary infrastructure (US pp). According to a study by the European Commission, the Union's share of energy from foreign sources is predicted to rise from about half in 2000 to two-thirds by 2020 (US pp). Due to environmental concerns and the phasing out of much of the EU's nuclear energy capacity, gas use will rapidly increase (US pp). Deliveries from the Yamal-Europe pipeline and the Blue Stream pipeline will increase Russia's gas sales to the EU and Turkey by more than forty percent over 2000 levels in the first decade of the twenty-first century (US pp). As a result, Russia's share of total European demand will rise from twenty-seven percent in 2000 to thirty-one percent in 2010 (US pp). Moreover, as the largest energy supplier outside of OPEC, Russia will be well positioned to marshal its oil and gas reserves to support domestic and foreign policy objectives (US pp). Algeria, which has the world's eighth largest gas reserves, is also seeking to increase its exports to Europe by fifty percent by the end of the decade (US pp).

On June 07 2005, Mikal E. Herberg, Director of the National Bureau of Asian Research Committee appeared before the Senate Foreign Relations Subcommittee on East Asian and Pacific Affairs (Herberg pp). According to Herberg, the issues emanating from China's growing energy needs are so important that NBR is organizing a conference for September 2005 in Washington, D.C. entitled, "China's Search for Energy Security and Implications for the U.S." (Herberg pp). Top energy and geopolitical experts will discuss a wide range of issues, including the outlook for

China's energy needs and energy imports, and its emerging and active energy security strategy (Herberg pp). Energy has become a central factor in shaping China's global geopolitical and diplomatic architecture in key oil and gas exporting countries and regions, such as the Persian Gulf, Central Asia, Russia, Africa, and more recently, the Western Hemisphere (Herberg pp). At present, energy nationalism is on the rise in Asia with ominous implications for Asia's future, as energy and strategic relations become increasingly intertwined (Herberg pp). China is the second largest energy consumer in the world, after the United States, and this booming energy demand growth is a reflection of its rapid economic and trade growth, urbanization, population growth and rising per-capita incomes (Herberg pp).

Other areas of Asia are also experiencing a period of extraordinary energy demand growth due to rapid economic growth and industrialization (Herberg pp).

The main difference between China and the rest of Asia is the sheer scale of China's energy demand due to the size of its economy and population, as well as the peculiarities of its domestic energy supply base (Herberg pp). This rapid demand growth is can be seen across the fuel spectrum including oil, natural gas, electricity, coal, nuclear and hydroelectric resources (Herberg pp).

Large domestic supplies of coal have dominated China's domestic energy use and continues to account for two-thirds of total energy consumption, however, rapid economic growth has accelerated oil demand growth and China's decision to expand the use of natural gas will boost future gas consumption (Herberg pp). Although China has been Asia's largest oil producer since the mid-1960's, oil demand is rapidly outrunning the country's domestic oil resources, resulting in rising oil imports (Herberg pp). China is now the third largest oil imported behind the United States and Japan, importing more than forty percent of its total oil needs (Herberg pp). According to the International Energy Agency, China's oil imports will rise more than five-fold by 2030, accounting for eighty percent of its total oil needs (Herberg pp). China, as the rest of Asia, will become heavily dependent on the Persian Gulf for future supplies, and will increasingly have to transit a series of maritime chokepoints (Herberg pp). The East-West Center forecasts that within the next ten years, seventy percent of China's oil imports will come from the Middle East (Herberg pp).

The demand for electricity has also accelerated in recent years, forcing China's government to seek fuels to generate more electricity, leading China to rely heavily on its largest domestic energy resource, coal (Herberg pp). China is the largest producer and consumer of coal in the world and coal

accounts for over eighty percent of electricity generation and accounts for two-thirds of China's total energy use (Herberg pp). Coal consumption is expected to double by 2025 with alarming environmental and health implications, as the country will account for one-quarter of the world's CO₂ emissions (Herberg pp). Moreover, it is expected that China will become a net importer of coal within the next ten years (Herberg pp).

This boom in electricity demand is also driving plans for the largest single country nuclear power building program in the world, as China plans to build two large nuclear plants per year over the next twenty years (Herberg pp). Extensive hydroelectric development is also planned, and policies are being developed to accelerate the use of renewables, such as solar and wind, however, these will make only a small dent in the electricity demand curve even under the most optimistic of forecasts (Herberg pp).

All of this raises a range of serious environmental and health concerns not only for China, but for the region and the United States (Herberg pp). Acid rain from China's coal burning is already a major problem in Northeast Asia and is causing diplomatic tensions with Japan and South Korea (Herberg pp). Moreover, there is evidence of mercury from China's coal burning being drafted by the jet-stream all the way to North America (Herberg pp). Rising coal consumption along with booming oil consumption will make China the largest source of carbon dioxide emissions globally which raises serious concerns about the effectiveness of any global effort to deal with controlling carbon emissions (Herberg pp).

Martha Caldwell Harris reports in "The Globalization of Energy Markets," that in this era of globalization, one nation's choices will affect the calculus of neighbors (Harris pp). When China courted Saudi Arabia with promises of assured imports, Japan was "rocked with the loss of the Arabian oil concession" (Harris pp). Asia's growing dependence on Middle East oil imports will create new imperatives to strengthen relationships with suppliers (Harris pp). Harris believes that new technologies can help address the environmental problems that are certain to grow more serious and the world population increases to nine billion in the coming decades (Harris pp). Fossil fuel use is the major cause of environmental problems, especially in developing nations where local and regional pollution is growing, and despite promises of hybrid cars and distributed energy generation, such as small turbines and decentralized power generation, market signals have not supported early commercialization (Harris pp).

Seeing global warming as a major threat, Japanese energy experts believe that as it becomes more apparent, there will be negative impacts on energy security (Harris pp). Their perceptions of energy security reflect a broader definition of risk and a greater focus on the Asian region, however even in Europe, there is renewed concern about energy security (Harris pp). According to recent forecasts, the overall import dependence of the European Union will rise to seventy percent for natural gas, eighty percent for coal, and ninety percent for oil by the year 2020 (Harris pp). Imports of Russian gas could reach as much as forty-five percent of the EU's total, and as energy demands in the developing world rises to surpass the demand of the Organization for Economic Cooperation and Development during this time frame, the EU share of global energy demand will shrink to roughly ten percent (Harris pp).

Also of major concern is the fact that the shallow Straits of Malacca and the sea-lanes between the Middle East and Asia will become more congested with tankers and other ships carrying fuel and commodities (Harris pp). Ninety percent of Japan's oil imports and the majority of South Korea's and Taiwan's oil imports flow through these waters (Harris pp). In fact, every day more than two hundred vessels pass through the Malacca, Sunda and Lombok Straits, and the South China Sea (Harris pp). More than \$1 trillion in international trade passed through these waters in 1994, which have seen an increase in serious accidents since the early 1990's, as well as an increase in piracy, kidnapping, and other acts of violence by non-state actors, such as left-wing rebels in the Philippines (Harris pp). The likelihood of even more accidents and acts of terrorism and piracy throughout the region in the future, has led some to call for a change in the transit passage law enshrined by the Law of the Sea separating commercial and military traffic (Harris pp). The objective would be increased regulation of commercial vessels in the Straits of Malacca to ensure navigation safety, because cleaning up after a major oil spill and relief efforts to deal with terrorism or piracy could be much more costly after the fact (Harris pp). Another way to address the vulnerabilities in energy transportation through the sea-lanes would be to develop regional emergency response mechanisms (Harris pp).

For years, Japanese firms have been mining coal in Australia, developing natural gas resources in Indonesia, and purchasing oil from China, and with greater openness come new possibilities (Harris pp). For example, Tokyo Electric Power has stakes in new power-generating ventures in Malaysia and Vietnam (Harris pp). Gas and

electric power are the focus of networks of growing international joint ventures that include firms from several Asian countries, as well as from the United States (Harris pp). These corporate linkages extend further and deeper into domestic economies, and in many cases, can stimulate market-oriented corporate restructuring and advanced technology development, and more importantly lead to new security challenges (Harris pp).

In 1996, almost one-fifth of Japan's natural gas came from Indonesia, a country where in many regions, violent independence movements have threatened central authority (Harris pp). Electric power, gas, and steel companies have long-term contracts of liquefied natural gas, LNG, imports from Indonesia that extend for more than a decade (Harris pp). Two-fifths of Indonesia's LNG exports come from Aceh, in the western end of Sumatra, that is overwhelmingly Islamic and the rural population resent the wealth of the Japanese who run the industrial enclave (Harris pp). The potential fragmentation of energy and resource rich regions poses major problems for the central government as well as for the importers whose investments become vulnerabilities (Harris pp). Japan, South Korea, Russia, China, Taiwan, India, and Pakistan all have commercial nuclear power programs, and four of these states have tested and/or developed nuclear weapons (Harris pp). Japan's nuclear power has been the central pillar of its energy policy, and is seen as the country's only hope for gaining a degree of autonomous control and for meeting environmental commitments (Harris pp). Thus, Asia has become the new center of gravity for the global nuclear industry, as additions to capacity in this region are projected to make up at least three-quarters of the world's total over the next twenty years (Harris pp). Therefore, for safety, environmental, and nonproliferation reasons, advanced technology cooperation in energy among Asian nations is essential (Harris pp).

In 1987, the term Sustainable Development first gained attention with a report from the World Commission for the Environment and Development what defined it as development "that meets the needs of the present without compromising the ability of future generations to meet the needs of others" (Young pp). The global mining industry finds survival economies less restrictive, because the local populations view a mine as an economic boost, for it mean potential jobs and the avoidance of poverty (Young pp). In other words, the local population is less concerned with potentially polluting activities so long as they have a way to earn a living and support their families (Young pp). In days past, when a mine became exhausted of resources, it was sealed and the mining company simply moved on to

the next community (Young pp). The SD movement tries to make the mining companies view the communities as places of long-term investment and ensure a health care and educational system, and that the local population has a means for survival (Young pp). This holds true for all industries that involve natural resources, whether oil, chemical, coal, etc. (Young pp). In 1998, nine of the world's largest mining companies initiated the Global Mining Initiative and commissioned a comprehensive study of the sustainable development challenges facing the mining industry (Young pp).

For example, Papa New Guinea is a developing country that has extensive mining activity, and among the negative consequences this country has experienced are health problems resulting from overexposure to mercury, pollution of fishing grounds, and the degradation of rivers and creeks (Young pp).

The positive consequences include investments in infrastructure, investments in local health and education, as well as improved attention to pollution controls (Young pp).

In "State of the World 2004 – Special Focus: the Consumer Society," published by the WorldWatch Institute, the authors give examples of how growing numbers of universities, corporation, government agencies and other institutions around the world are incorporating environmentally friendly, so-called "green" concerns into their purchasing habits (Rutsch pp). This green purchasing power can play a key role in supporting changes toward an environmentally sustainable world (Rutsch pp). With respect to sustainable energy consumption, people in countries like Norway and Japan enjoy a high standard of living while using less energy per person than the average American (Rutsch pp). The authors stress that government policies, including regulation, standards, subsidies and taxes, are "critical for improvements in energy efficiency and conservation" and for the sustained growth of "cleaner and greener" energy technologies (Rutsch pp). However, individual consumers can play a key role through their daily choices "by creating demand for products and services that are more energy-efficient and by influencing wider policy decisions" (Rutsch pp). It is due to extreme consumption that humanity and the earth have come to the edge of an environmental abyss, as the world's finite resources are rapidly being depleted and degraded, and although the use of resources and generations of pollution and wastes continue to grow, there is still a chance to control consumption attitudes with coalitions at the government, business, and consumer levels (Rutsch pp).

In the face of these environmental dangers, many believe that new regulations should be enacted both locally and globally (Marber pp). Increased environmental awareness among wealthier nations may lead to domestic policies that will raise costs to businesses and consumers, which in turn could curb economic expansion (Marber pp). Peter Marber, in his article, ‘Globalization and its Contents,’ published in *World Policy Journal*, says that one step in the right direction would be increased public spending on alternative and renewable energy sources in the wealthier countries (Marber pp). The world is clearly under-powered, and the need for diversified energy is growing every day, therefore the benefits of a burgeoning alternative energy sector could be “multiplicative” (Marber pp). Marber states:

First, it might spur new economic growth areas for employment in rich countries, supplying them with potential technologies for export while reducing their reliance on foreign oil. Second, it might encourage developing countries that are over-reliant on oil exports to develop and modernize their economies and societies.

Third, it would allow developing countries to build their infrastructures with a more diversified, sustainable energy approach than the first wave of industrializing countries (Marber pp).

Globalization and its major engines of burgeoning human capital, freer markets, and increasing cross-border interaction, have created a New World order that has incited passionate debate, pro and con (Marber pp). Marber points out that even with its positive trends, globalization is not a perfect process, and is not a panacea for every problem for every person at every moment in time (Marber pp). In reality, it is a messy, complicated web of interdependent relationships, some long-term, and some fleeting (Marber pp)

In May 2005, Saudi Oil Minister, Ali Naimi, spoke before the World Affairs Council of Northern California and Council on Foreign Relations, stating that he believe globalization held the promise of a better way of life for the world’s population, but realized that it will not be easy (Naimi pp). The world will be faced with tradeoffs in an effort to balance economic growth, quality of life, the environment, culture and tradition (Naimi pp). In promising the benefits of globalization, it is important to understand the essential role of energy, for economic activity requires energy to produce goods, move them to markets and sell them to consumers (Naimi pp). Moreover, energy provides the world with many of the conveniences of modern life,

and without energy, economic progress is not possible (Naimi pp).

Due to globalization, the world’s demand for energy will only continue to grow, and given the state of technology, there are currently no viable substitutes for oil, especially in the transportation sector where oil accounts for ninety-five percent of the energy consumed globally (Naimi pp). Therefore, it is almost certain that oil will remain the fuel of choice in transportation, both from the standpoints of economics and ease of use, for at least the next thirty years (Naimi pp). Although there are alternative technologies, such as fuel cells and battery-powered vehicles, that hold promise for the future, in reality, none are currently close to being commercially competitive with gasoline and diesel powered vehicles based on the internal combustion engine (Naimi pp). However, some new fuel-saving technologies, like hybrid vehicles, are commercial today and are becoming increasingly competitive in the market place, and demonstrate how technological advances can enable the world to more efficiently utilize oil resources and minimize the impact on the environment (Naimi pp).

In this era of globalization, oil has become a financial investment asset, similar to currencies, equities and bonds (Naimi pp). Oil futures and over-the-counter markets are now attracting huge sums of money from hedge funds and institutional investors seeking to maximize returns (Naimi pp). These investment decisions are not necessarily based on prevailing market fundamentals, but rather on expected returns relative to alternative investments (Naimi pp). Thus, it is more difficult to stabilize markets due to the massive funds involved (Naimi pp). Globalization will expand the world’s economy when in turn will create an unprecedented demand for additional oil (Naimi pp). While demand growth is expected to be large, Naimi believes the remaining oil resources are significant, and that technological innovations will provide the key to achieving balance markets by improving the efficiency of producing and consuming oil (Naimi pp). Although both producers and consumers benefit from stable and predictable prices, he cautions to be mindful that oil’s new role as an investment asset and the lack of market transparency complicate the task of achieving stability (Naimi pp).

As globalization continues, free markets will ensure that investment capital and fossil fuels are distributed efficiently, however, the future could involve more regulation and confrontation (Globalization pp). In other words, rather than free markets, anxious governments will decide how capital and energy supplies are apportioned, “who gets how much access to the coveted resource rich areas will

be affected by geopolitical competition or even conflict” (Globalization pp). Thus, rather than globalization, this would be “deglobalization” (Globalization pp).

Massive industrialization in Asia and the continued high industrial output of advanced economies pose major global environmental concerns (Laducina pp). Energy demands are only going to skyrocket as populations and industrial outputs continue to grow (Laducina pp). Although proven oil reserves will likely satisfy demand for the next several decades, the largest reserves are in the Middle East and Central Asia, where chronic political instability continually threatens supply (Laducina pp). Europe and the United States are such large consumers of natural gas that domestic reserves are dwindling and utility prices are rising (Laducina pp). China and India combined will account for two-thirds of global demand for coal through 2030, and China alone is already the world’s second largest oil importer (Laducina pp). As key resources dwindle, prices rise, and supply shrinks, competition for these resources will grow intense, with conflicts over water and oil the most likely (Laducina pp). As energy costs and environmental concerns grow however, alternative sources will become price competitive, just as solar power is on Long Island, New York, and wind power is in Minnesota and elsewhere (Laducina pp).

Developing countries entering the marketplace not only need more energy to produce and move their products, but they expect the same perks, such as modern appliances, etc., as Western societies, and that means increased energy consumption and increased energy demand. The genie is out of the bottle. Globalization is a matter of fact, and no one seems to have any quick fixes or even a viable prognosis concerning the inevitable dilemma the world faces as natural resources are depleted amid a continuing and increasing demand for more and more energy. It appears globalization is a “catch 22.”

References

- “Globalization and Energy Supply: Strategic Risk in the 21st Century.” A Deloitte Research Viewpoint. May 2004. <http://www.deloitte.com/dtt/research/0,1015,sid%253D1009%2526cid%253D50472,00.html>
- Harris, Martha Caldwell. “The Globalization of Energy Markets.” http://64.233.187.104/search?q=cache:GwIxkr0pEwYJ:www.ndu.edu/inss/books/Books_2001/Global%2520Century%2520-%2520June%25202001/C12Harri.pdf+globalization+and+the+demand+for+energy&hl=en
- Herberg, Mikal E. “Consequences of a Growing China.” Congressional Testimony. 6/7/2005. eLibrary. (accessed June 17 2005).
- Laducina, Paul A. “Managing global risk to seize competitive advantage.(analysis of risk management).” Ivey Business Journal Online. 3/1/2005. eLibrary. (accessed June 17 2005),
- Marber, Peter. “Globalization and its contents.” World Policy Journal. 12/22/2004. eLibrary. (accessed June 17 2005).
- Naimi, Ali. “Globalization and The Future of the Oil Market.” Middle East Economic Survey: VOL. XLVIII No 22 May 30 2005. (accessed June 17 2005). <http://www.mees.com/postedarticles/oped/v48n22-5OD01.htm>
- Rutsch, Horst. “State of the World 2004; Special Focus: The Consumer Society.” UN Chronicle. 12/1/2004. eLibrary. (accessed June 17 2005).
- “The US National Intelligence Council on the changing geopolitical landscape.” Population and Development Review. 3/1/2005. eLibrary. (accessed June 17 2005).
- Young, Scott T. “Leading environmental change: the case of the global mining industry.” Review of Business. 1/1/2005. eLibrary. (accessed June 17 2005).

Overview and Analysis on Cyber Terrorism

Steve Saint-Claire

MSc and candidate to PhD in Computing Technology by the Department of Engineering and Technology of the School of Doctoral Studies of the EU

[Email: s.saintclaire@det.sds.eu](mailto:s.saintclaire@det.sds.eu)

Abstract

There are many forms of terrorism on the Internet. Some are not dangerous enough to be deemed a simple spread of information instead of terrorism. They are simple show of skill and are harmless. Acts of cyber-crimes may involve stealing of money, company secrets, or attacking country's infrastructure and causing real damage. Cyber terrorism is an impending threat to the United States, or any other technologically advanced country. Even nations with more primitive technology can be negatively affected by the "ripple effect". With the excess of technology increasing at a tremendous rate, the threat of cyber terrorism will only get worse. This article provides analysis on definition, methodologies, participants and different forms of protection against cyber terrorism. **Keywords:** Cyber Terrorism.

It has been called "information war" or "cyber terrorism" the capability to let loose terror and destruction with a few well-aimed clicks on a computer keyboard. Although it is an unusual and a potentially lethal subject, the public does not know very much about it. However, experts in and around the country, to the top level of the federal government, are taking it seriously.

The FBI defines terrorism as the unlawful use of force or violence against persons or property to intimidate or coerce a government, the civilian population, or any segment thereof, in furtherance of political or social objectives. Cyber-terrorism can be the use of computing assets to threaten or force others. An example of cyber-terrorism could be hacking into a hospital computer system and changing someone's medicine prescription to a lethal dosage as an act of revenge. It sounds far fetched, but these things can and do happen. Cyber terrorism has also been defined as "the use of computing resources against persons or property to intimidate or coerce a government, the civilian population, or any segment thereof, in furtherance of political or social objectives."¹.

Cyber Terrorism

There are many forms of terrorism on the Internet. Some are not dangerous enough to be deemed a simple spread of information instead of terrorism. They are simple show of skill and are harmless. Acts of cyber-crimes may involve stealing of money, company secrets, or attacking country's infrastructure and causing real damage.

In mid 1990's Zapatista guerrillas used the Internet to rally international pressure on the Mexican government to end its assaults against them. They used the Internet to extend rumors of rebel victories. In other words, they accomplished more with a laptop and an Internet connection than they could with 12000 armed troops.

Historically, American society has been faced with a variety of threats: racial tension, depression, segregation, terrorism, and computer glitches to name a few. Though many can now be considered "history," these threats have frequently received extensive public and media attention. On the other hand, the appearance of the "computer age" has produced a mutation of an already well-known and much feared threat, Cyber terrorism. Unluckily, this new menace has not fully gained the attention of a more informed and educated public.

A difference should also be made between terrorists who make use of available technology, and the pure "cyber-terrorist." Traditional terrorists may increase their arsenal of more conservative methods, such as bombings, hijackings, and murders, with new methods such as computer viruses, Radio Frequency Weapons, , and "denial of service" attacks. However, the pure cyber-terrorist may also exist. This variety of terrorist may do without the conventional approach of terrorism, instead exploiting computer technology to put into effect demands, gain ransoms, or generally cause destruction upon the world population. additionally, the cyber-terrorist can achieve these objectives without exposing himself to actual harm. Although the pure cyber-terrorist has not appeared in the news yet, current

hacker and denial of service attacks may provide a peek as to what the pure cyber-terrorist can accomplish.

The largest threat associated with cyber terrorism and the biggest advantage to the pure cyber-terrorist) is the "detached" sort of attacks. Borders are not crossed, bombs are not smuggled and placed, hostages are not captured, and terrorists do not surrender their lives. Openly stated, "tomorrow's terrorist may be able to do more with a keyboard than with a bomb."

Cyber terrorism is an impending threat to the United States, or any other technologically advanced country. Even nations with more primitive technology can be negatively affected by the "ripple effect". With the excess of technology increasing at an tremendous rate, the threat of cyber terrorism will only get worse.

Examples of Cyber-Terrorism

By using the Internet, the terrorist can influence much wider harm or change to a country than one could by killing some people. From immobilizing a countries military defense to shutting off the power in a large area, the terrorist can have an effect on more people at less danger to him or herself, than through other means.

Cyber terrorism takes many forms; following are some examples of cyber-terrorism in its many forms:

1. Cyber-terrorists often commit acts of terrorism purely for private gain. Such a group, known as the Chaos Computer Club, was exposed in 1997. They had formed an Active X Control for the account. Without difficulty, this could be used to steal money from users all over the world that have the Quicken software installed on their computer. This type of file is only one of thousands of types of viruses that can do everything from simply annoy users, to disable large networks, which can have disastrous, even life and death, results.
2. Cyber-terrorist is interested in gaining publicity in any probable way. For example, information warfare procedures like Trojan horse viruses and network worms are repeatedly used to not only do harm to computing resources, but also as a way for the designer of the viruses to "brag." This is a serious moral issue because many people are affected by these cases. The viruses can use system resources until networks become ineffective, costing companies lots of time and money. In addition, depending on the type of work done on the affected computers, the damage to the recipients of that work could be lethal; it could have unpredictable effects that could have dreadful consequences.

3. In one of its more remarkable forms, cyber-terrorism can be used for a murder. In one case, a mafia boss was shot but survived the shooting. That night while he was in the hospital, the assassins hacked into the hospital computer and altered his medicine so that he would be given a lethal injection. He died a few hours later. Then they changed the prescription order back to its accurate form, after it had been incorrectly dispensed, to cover their tracks so that the nurse would be blamed for the "accident".
4. Terrorism can also come in the shape of misinformation. Terrorists can repeatedly say what they please without fear of retaliation from authorities or of answerability for what they say. The rumor that a group of people was stealing people's kidneys for sale was spread via the Internet. The report unnerved thousands of people.
5. Small strikes come in the form of "data diddling", where information in the computer is distorted. This may involve altering medical or financial account or stealing of passwords. Hackers may even prevent users who should have access from gaining access to the machine help because the computer would not allow the necessary access for the doctor to save his or her life.

Cyber-Terrorist Weapons, Methods, and Techniques

Cyber terrorism flourishes on the development of new technologies. Although often developed for some other non-threatening purpose, new technological developments grant terrorists and cyber-terrorists with new weapons for their arsenal. The Internet may also aid in the distribution of technical data related to these new technologies.

Additionally, quite a few grave dangers are presented by many of these new weapons, they may allow the user to attack from a great distance; Moreover, as the most technologically reliant advanced nation, the United States is the most vulnerable nation on earth to cyber-attack.

Worms

Recently, the invention of an unprecedented number of prolific worms, (e.g. Code Red, Ramen, Lion) some of which are suspected of having been created in response to political events. The weaknesses worms utilize, are usually well known to system administrators and able to be remedied, but often go un-patched on enough systems to cause major problems in the information infrastructure.

A worm similar to Code Red could do much more severe harm with only minor design modifications. This analysis points to the conclusion that if maximum destruction is a hostile challenger's goal, worms are a cost effective way to extensively disrupt the United States National information Infrastructure.

The terms virus and worm are regularly used synonymously to describe malevolent, autonomous computer programs. Most current computer viruses are in fact worms. The worm epidemic of recent months, enabled by a common buffer overflow exploit, demonstrate this experience. Buffer overflows allow attackers to hijack legal computer programs for illegal purposes, and they were once the dominion of only the most selected programmers. In the past five years, however, buffer overflow attacks have become increasingly accepted, and they are now the favorite among hackers of all skill levels. In June 2001, a computer security company identified a weakness in a popular web server program that could lead to a buffer overflow exploit.² The company published a benign exploit to reveal its point, but within days of the initial report a malicious program exploiting the identified weakness was making the rounds in the hacker world.

Unauthorized Intrusions

Unauthorized computer infringement and the loss of sensitive information are of great alarm to businesses and governments alike. Although there was a time when intrusions were limited to curious hackers, organized crime and other organized groups eventually realized the profit of collecting poorly protected electronic information for financial or other gain. In March 2001, the NIPC issued a warning that organized crime had made noteworthy inroads in cyberspace.³ A series of intrusions, collectively known as Moonlight Maze, in U.S. Government systems over a period of several years may have instigated in Russia. The first attacks were detected in March 1998 and, in the course of this continued assault, hundreds of unclassified networks used by the Pentagon, the Department of Energy, NASA, as well as a range of defense contractors, were compromised. While authorities insist that no classified systems were violated, it is acknowledged that vast quantities of technical defense research were illegally downloaded.

Domain Name Service (DNS) Attacks

Computers connected to the Internet communicate with one another using numerical IP addresses. Domain name servers (DNS) are the .Yellow Pages. that computers consult in order to obtain the mapping between the name of a system (or website) and the numerical address of that system. If the DNS server supply an erroneous numerical address for the web site, the user's system would connect to the incorrect server. Making matters worse, this fake connection would likely be completed without arousing the user's doubt. The result would be that the user is offered a web page that he believes is on the desired web server but, in reality, is on the attacker's server. An attacker could distribute fake information with a triumphant attack on a select DNS server (or group of servers), bypassing the need to break into the actual web servers themselves. Moreover, a DNS attack would prevent access to the original web site, depriving the site of traffic.

Distributed Denial of Service (DDoS) Attacks

Distributed Denial of Service (DDoS) attacks have also evolved over time. DDoS attacks utilize armies of .zombie machines taken over and controlled by a single master to overpower the resources of victims with floods of packets. These attacks are best known in the context of the high-profile attacks of February 2000, where popular e-commerce web sites were shut down by simultaneous attacks. Since then, the reputation of high-speed home Internet access has increased, and the commanders of DDoS zombie armies are taking advantage of this popularity. Preying on the careless security of the average home computer user, attackers have found ways to plant malicious programs to give themselves remote control of home computers. Many of these machines are now unwitting participants in DDoS attacks.⁴

The Potential "Players" in the World of Cyber-Terrorism

Cyber terrorism has the ability to modify the face of modern terrorism, as we know it. For the most part, terrorist organizations inhabit isolated areas of the world, only venturing out when the time is right to strike.

Because of this lifestyle, terrorist groups are restricted in the means through which they may broadcast their messages and ideals to the world audience. Usually, the most successful way to get press is to terrorize or actually carry out an act of terrorism. However, this requires the terrorists to expose themselves. It does not require the expertise of a

rocket scientist to realize the inherent danger in this method of operation. The ability to transmit an unlimited amount of information from the comfort of your own tent, cave, bunker, or palace. One could even electronically organize plans or carry out acts of terrorism with little more than a computer and an Internet connection.

In age of the information upheaval, terrorist organizations, which usually have no admission to television or radio communications, can easily broadcast their messages to entire world via the Internet. In fact, many of the major terrorist groups either maintain their own website, or have sites dedicated to them. Aum Shinrikyo, the group who orchestrated the gas bombing on Tokyo subways, operates its own site. Hizballah, has a site detailing its mission. These sites allow terrorist organizations to contact eventual target audience- the worldwide population.

Dangerous prospective lie in wait, in the capability of terrorists to access and exploit a worldwide soapbox. theoretically speaking, even if only two out of every 200 persons who visit a site are influenced by the messages, this number can rapidly grow into thousands as more persons around the world use the Internet as a source of information.

The not-so-hypothetical pure cyber-terrorist may also use the Internet as the only means for carry out terrorist attacks against civic utilities, national governments, or a assortment of other targets.

It is clear that terrorist organizations are starting to use the "information superhighway." It is almost definite that terrorist organization will go on to migrate to the Internet at an ever-increasing rate.

Terrorist Groups

It is uncertain whether Osama bin Laden's international Al Qaeda organization or other terrorist groups have developed cyber warfare means, or how widespread these capabilities may be. To date, few terrorist groups have used cyber attacks as a weapon.

However, terrorists are known to be extensively using information technology and the Internet to devise plans, raise funds, spread misinformation, and communicate securely.⁵ For instance, the convicted terrorist, Ramzi Yousef, who was accountable for planning the first World Trade Center bombing in 1993, had details of future terrorist plots (including the planned bombing of 12 airliners in the Pacific) stored on encrypted files on his laptop computer. At the same time, the September 11, 2001 attacks on the World Trade Center and Pentagon and earlier terrorist targets, such as the British security forces detection

that the Irish Republican Army (IRA) planned to destroy power stations around London, reveal an increasing wish by terrorist groups to attack critical infrastructure targets. The World Trade Center attacks not only took lives and property but also closed markets and ruined an important part of the financial information infrastructure in New York City. Thus, trends seem clearly to point to the possibility of terrorists using information technology as a weapon against critical infrastructure targets.

Terrorist Sympathizers and Anti-U.S. Hackers

If historical drift continue, attacks by those sympathetic to the terrorist group(s) responsible for the September 11, 2001 attacks on the United States and those with general anti-U.S. and anti-allied sentiments are more likely than attacks by the terrorists themselves or by nation-states. If the American campaign against terrorism is perceived as a crusade.

Thrill Seekers

Any conflict that plays out in cyberspace will invariably attract a huge number of hackers who merely want to gain notoriety through high profile attacks. This group of attackers may not be driven by political or ideological zeal, but simply the desire to attain bragging rights about their exploits.

White Supremist and Hate Groups

The cost of right-wing extremism on the Internet should be of concern to all. Online, racists, anti-Semites, and anti-government extremists can contact a much larger audience than ever before. Anyone using the Internet may involuntarily be bared to hate on the web. While estimating the affect of extremist material on the vast population of Internet users presents huge problems, significant information does exist about a small subset of this group: right-wing extremists. In the Internet age, these extremists can communicate with thousands of their compatriots with the click of a mouse. The Internet has offered the far-right fringe with formerly inconceivable opportunities.

When uninformed or easily influenced people – particularly children – come across hate propaganda, they can fall victim to its misleading logic and adopt hateful beliefs themselves, occasionally going so far as to act on what they have read.

No reliable capacity has been taken of the number of Internet users who find and read hateful material online. Nor can one safely simplify about the ways in which this material affects the viewpoint and actions of those who read it. Extremist readers can learn about marketing frauds, or discover how to use "paper terrorism" techniques such as filing specious liens.

Bomb Making

Directions for making bombs and other terrorist tools are readily available online to all types of extremists, and many white supremacist Web sites have either posted bomb-making instructions or linked to such material.

"In 1999, British neo-Nazi David Copeland planted nail bombs in a Black neighborhood, an Indian area, and a gay pub in London, killing three and injuring more than a hundred. Copeland later wrote, "I bombed the blacks, Paki's [sic], [and] Degenerates," and he boasted, "I would of [sic] bombed the Jews as well if I got a chance." A court handed Copeland six life sentences for his crimes. He had learned how to build his bombs by visiting a cyber cafe, where he downloaded *The Terrorist Handbook* and *How to Make Bombs: Book Two* from the Internet.

"Explosives are very effective in our cause," writes "Death Dealer," the anonymous creator of the racist skinhead site *Better than Auschwitz*. "They should be deployed more." *Better than Auschwitz* includes pictures of bombing victims and detailed bomb making instructions."⁶

Directions for Weapons Use

A "Nigger Baiting Made Easy" section explains "the various methods of selecting muds and queers, and getting them to fight, or throw the first punch." Such material resembles the instructions White Aryan Resistance gave the skinheads of East Side White Pride before their violent rampage in Portland.

Supplying Hit Lists

Online, radicals may find direction not only on how to attack, but also who to attack. Anti-government sites often post information about judges, law enforcement officers, and other government officials. next to graphics dripping with blood and links to sites calling the murder of abortion providers "justifiable," the Nuremberg Files

Web site supplies detailed personal information about doctors who supposedly provide abortions, including their social security numbers, license plate numbers, and home addresses. The list of doctors reads like a list of objects for killing.

Freedom of Speech and Censoring the Net

Today, the Internet is susceptible to attack. The attacks are hard to avert and react to, and the person responsible often go unidentified. The Internet can be abused and misrepresented, to the loss of all legitimate users and only to the benefit of those with malicious intent. Without steps to make the Internet a safer and more consistent atmosphere, the operation of our critical Infrastructures will remain at risk.

Governments are appreciating the need to protect their information and critical infrastructures in reply to these threats and are responding consequently. Some governments distinguish that it is not adequate to address only the local or national aspects of safeguarding information and critical infrastructures. Because attacks against the Internet typically do not require the attacker to actually be present at the site of the attack, the risk of being recognized is drastically reduced.

Besides the technological challenges this presents, the legal issues involved in tracking and prosecuting intruders adds a layer of difficulty as they cross multiple geographical and legal boundaries. An effective answer can only come in the form of international collaboration.

In the United States, for example, Presidential Decision Directive 63 (a white paper on critical infrastructure protection) states, "Addressing these vulnerabilities will necessarily require flexible, evolutionary approaches that span both the public and private sectors, and protect both domestic and international security.... The Federal Government shall encourage international cooperation to help manage this increasingly global problem."

In the area of law enforcement, the Internet constitutes a new patrol area in many respects. Unlike jurisdictions based on national and political borders, the digital information infrastructure does not have a central location in the physical world. So not only is take action to attacks tricky technically but also many of the established methods for practicing law enforcement are ineffective. Trying to tackle the problems in one group without input and feedback from the others is likely to result in defective or partial solutions. U.S. government legislation (the Digital Millennium Act) ensuing from the World Intellectual Property Organization

(WIPO) treaty resulted in a bout of panic throughout the Internet security community. Practitioners, researchers, software vendors, and incident response teams realized that aspects of their work that address security flaws to reduce risk to our critical infrastructures might become illegal under the proposed legislation. This was clearly not the original intent of the treaty or the resulting legislation. This is just one example of the pressing need for continuing communication among policymakers, technologists, and others to make certain that future policies and settlement on a national and international scale are realistic and effective.

Information exchange and interaction among many parties is paramount to producing complete and practical approaches and solutions to the complex problems faced. Simply bringing parties together is not enough.

Present Difficulties

Many network protocols that now form part of the information communications were designed without computer security in mind. Without a secure infrastructure, it is difficult to avoid security troubles and settle computer security occurrences. The combination of fast changing technology, rapidly increasing use, and new, often unimagined uses of the information infrastructure gives rise to a unpredictable situation in which the nature of coercion and vulnerabilities is difficult to appraise and even more difficult to predict.

It is cheap (the cost of a personal computer and Internet access), speedy (less than a minute), straightforward (using freely available intruder tools) for anyone to initiate attacks against critical information infrastructures. Conversely, it is costly (international effort and funding), long-term (research, development, and deployment), and intricate (technically and politically) to take the steps needed to harden the information infrastructure to make it less vulnerable to attack, and to facilitate in responding more effectively and efficiently when attacks do happen.

In general, incident response and computer security teams consist of professionals and technologists who have a wealth of operational experience but lack authority to make policy and security judgment in their organizations. They also may have limited funding and lack professional recognition.

American Society and Cyber Terrorism

In the last 50 years, the world has entered the information age when it comes to the way live, the way we work and

the way we play. Today, information and technology virtually rules our lives. Some people would dispute that our society depends excessively much on technology and that has become our weakness. Never before could a small and fanatical group of people with malevolent objectives and the right knowledge cause so much harm to a large and hence vulnerable entity such as a country government.

Currently there are no perfect ways to protect a system. The completely secure system can never be accessed by anyone. Most of the militaries classified information is kept on machines with no outside connection, as a form of prevention of cyber terrorism. Apart from such separation, the most widespread method of protection is encryption. The prevalent use of encryption is inhibited by the governments ban on its exportation, so intercontinental communication is left relatively insecure. The Clinton administration and the FBI oppose the export of encryption in favor of a system whereby the government can gain the key to an encrypted system after gaining a court order to do so. The director of the FBI's stance is that the Internet was not intended to go un-policed and that the police need to protect people's privacy and public-safety rights there. Encryption's drawback is that it does not protect the entire system, an attack designed to cripple the whole system, such as a virus, is unaffected by encryption.

Others promote the use of firewalls to monitor all communications to a system, including e-mail messages, which may carry logic bombs. Firewall is a relatively common term for methods of filtering access to a network. They may come in the form of a computer, router other communications device or in the form of a network configuration. Firewalls assist to define the services and access that are permitted to each user. One method is to screen user requirements to check if they come from a previously defined domain or Internet Protocol (IP) address. Another method is to prohibit Telnet access into the system.

Moral Concerns

The ethical concerns involved in cyber-terrorism are diverse. Any sort of crime or ethical violation can occur using a computer. Extortion of banks takes money from the banks, as well as their customers. The bank's, on the other hand, which many times decline to admit to their insufficient resistance violate the public trust that the bank will be secure. The unlawful changing medical records is unethical, as it can quickly and easily cause harm to another. Spreading disinformation is unethical in its lack of regard for the truth, as well as for the safety of and cost on others

who believe the misinformation. changing, damaging, or stealing others data is a violation of their privacy. The ordinary hacker is guilty of lack of regard for the privacy of the peoples systems that he or she would enter. Hacking-for-hire is additionally illicit because they openly sell their services to break into others systems.

Just as the real world terrorist acts have forced our society to become more watchful at airports, public places and such, cyber terrorism has enforced business and individuals to change the way they work and communicate with each other.

The effects of cyber terrorism are very much felt in today's society. Because these attacks can be seen not only in foreign far-away countries but also in our own "backyards" (places where we do business, our jobs, our homes), there is a sense of anxiety and concern created. Rather than seeing computers and the Internet as the unrealistic safe resource/ communication/ business tool it was created to be, it has grown to be an effective tool, which we need but against which we must also guard ourselves. Out of this insecurity, society has built up a mechanism to help deal and protect us from this reality. In the last ten years, we have seen innumerable books, articles, movies, headlines, investigative reports, and even whole sectors of the computing industry based on the anxiety & fear of the undesirable and unfamiliar. The effects are so deep that now words such as "computer virus" and "hacker" are now a common part of the English language.

Apart from the emotional and psychological cost to society, there is the economical effect. There have been reported instances where either individuals or institutions have paid millions of dollars to cyber-terrorist as ransom for threats made against them. Take for example the following statistics out of Great Britain.

"According to a source in Great Britain, terrorists have gained at least up to 400 million pounds from 1993 to 1995 by threatening institutions. Over the three years, there were 40 reported threats made to banks in the U.S. and Britain. In January of 1993, three separate incidents took place in London. During the sixth, a brokerage house paid out 10 million pounds after receiving a threat and one of their machines crashed. On the fourteenth incident, a blue-chip bank paid blackmailers 12.5 million pounds after receiving threats. Another brokerage house paid out 10 million pounds on the twenty-ninth incident...A Russian hacker, for example, tapped into Citibank's funds transfer system and took \$10 million." ⁷

Just alone from the occurrences mentioned in the above example, £42.5 million was paid out in only four out of the 40 reported cases. What about the other 36 cases? How about the **unreported** cases? How much was paid out for them? As you can deduce, the actual amount paid during that period for all CT cyber terrorism attacks could very well double or even triple the figure mentioned and these statistics are only from two countries. Think of what the amount can be worldwide.

Again, the effects are deeper than what one thinks. Just as the cost of a petty crime such as shoplifting eventually is pass on to the customer, so is the cost of cyber terrorism. The cost of replacing the lost funds, money spent in implementing a more secure environment, loss of revenue due to a smaller market share resulting from negative press, and other expenses are transferred over to the customer based in the form of higher fees for services and products. If the loss is significant enough, a company can fail and the effects are even greater due to the loss of jobs and similar consequences. If you are still not convinced, consider the effects of the opening scenario. What would happen if you lost e-mail for a day? How about not being able to make a phone call? Combine an Internet disabling attack with a strike against any of the nation's infrastructure systems (e.g. telephone, energy, transportation, etc.) and the results are devastating. When attacks reach this level of magnitude they can be classified as cyber-warfare.⁸

How to Protect Ourselves

Usually, there are several ways to protect ourselves. First, people must consider that cyber terrorism is valid and not only meant for government and large Corporations. Even if they were likely to be bigger targets for the malicious hackers, they are not the only targets. the next step is to protect your site with the range of apparatus available in the market place. If we take this a step further, one should learn to be proactive in protecting one's network. It makes life much straightforward when trying to prevent an attack from taking place as opposed to finding an attack, study the damage, and then prevent it from happening again. Being proactive requires a necessary step in Information security, training.

Finally, once a security system and policy is in place, the next step is caution. Among the objects that involve vigilance are the review of daily logs, performance of software updates, analysis of the latest hacker attacks and tools, & periodic revision of the security systems.

In his article "Cyber terrorism, information warfare, and attacks being launched now and in the future in the heartland of America" ⁹, Dave Pettinari lists the following list of to help guide us in protecting our networks:

- A security model that separates information in three categories: critical, sensitive, and public domain; with appropriate hardware/software security for each level.
- Strong passwords.
- Physical protection of good hardware and protected cabling.
- Firewalls between the network and outside.
- A good security policy with employees trained to appreciate its finer points.
- Audit trails for logins, operation of files, and successful/unsuccessful accesses (lots of attacks come from within).
- An intrusion detection system to identify harmful or malicious activity.
- Encryption to prevention interception of vital information.
- Workstations without disk drives and control/accountability of floppy disks to prevent evaporation of extremely sensitive information.
- Dial-up and Internet-access restrictions (For example, three attempts to get into our system and you are locked out prevents war dialing programs from hammering their way in).
- Background checks on personnel.
- Technical training and security awareness for all staff.

If this list is examined more thoroughly one will notice that it displays many of the best procedures that an information security specialist should follow.

Actions of the US Government to Fight Cyber Terrorism

Technologically advanced USA cannot deter cyber attacks. In fact, it could be said it makes an striking target. During the latest Yugoslavia conflict, Hackers with Chinese Internet addresses launched coordinated cyber attacks against the United States and allied forces.

US have created its own cyber warriors to use in offensive cyber attacks against Serbian computers and command and control systems. One can be certain that soon enough military will embrace information warfare as an effective way to combat an enemy wreaking serious damage to enemy infrastructure, without large expenses in war machinery and human lives.

The chance of a small group attacking the US government electronically is much bigger than the chance of an entire nation attacking US. US have already taken steps in order to prevent cyber crimes. This legislation plans to persuade businesses and government agencies to exchange information with the law enforcement agencies about security breaches. A more controversial legislation would permit US law enforcement agencies to read electronically encrypted mail and messages, under the same situations that they are allowed to tap phones. The private sector has also started to protect itself by forming organizations like Computer Emergency Response Team or CERT, which deals with system security incident response.

The US government has also taken more sweeping steps in preparing for cyber war. The National Security Agency hired its own hackers who managed to get root level access to 36 Department of Defense networks. This demonstrates that the US definitely is susceptible to cyber terrorism.

Finding answers to cyber-security weakness and attacks has been historically reactive. Attacks happen, analysis prepared and a patch provided, if possible. Individual attacks cannot continue to be solved on a case-by-case basis, and ignore the larger problem. A superior tactic is to apply practices and policies that develop the protection of our networks by upsetting a higher percentage of attacks. In other words, becoming more proactive in our attitude to cyber-security. By encouraging practices currently in place for more security-focused companies and tailoring them for other sectors,. Many companies, especially medium-sized and smaller firms are susceptible and looking for support in determining what security practices can help them better protect their systems.

Private and Public Cooperation

The safety and survivability of the Internet depends on the cooperation between the private and public sectors. Congress should encourage interaction between government and the private sector and should address problems such as exemption from FOIA and anti-trust barriers. In addition, Congress can set a great model for the private sector by escalating the security of all government systems, which historically have been out-dated and not met minimal standards for security.

The Internet Security Alliance is able to act as a bridge between the private sector and public sector by promoting best procedures and suitable data sharing mechanisms. The Internet Security Alliance is also involved in the following activities:

- Providing thought leadership on information security issues.
- Representing industry's interest on information security issues before legislators and regulators.
- Creating mechanisms that cause rapid development and implementation of information security practices, policies and technologies
- Identifying and standardizing best practices in Internet security and network survivability
- Creating a collaborative environment to develop and implement information security solutions
- Promoting universal sharing of information and intelligence on emerging threats/vulnerabilities/countermeasures

In reply to heightened awareness of the potential for cyber-terrorism President Clinton, in 1996, created the Commission of Critical Infrastructure Protection. The board found that the combination of electricity, communications and computers are essential to the continued existence of the U.S., all of which can be threatened by cyber-warfare. The resources to launch a cyber attack are commonplace in the world; a computer and a connection to the Internet are all that is really needed to cause disorder. Adding to the problem is that the public and private sectors are rather uninformed of just how much their lives depend on computers as well as the weaknesses of those computers. Another problem with cyber crime is that the crime must be solved, (i.e. who were the perpetrators and where were they when they attacked you) before it can be determined who has the real power to investigate the crime. The board recommends that significant systems should be isolated from outside connection or sheltered by sufficient firewalls, use best procedures for password control and protection, and use protected action logs.

Nearly all other government organizations have also formed some type of group to deal with cyber-terrorists. The CIA created its own group, the Information Warfare Center, staffed with 1,000 people and a 24-hour response team. The FBI inspects hackers and similar cases. The Secret Service tracks banking, fraud and wiretapping cases. The Air Force created its own group, Electronic Security Engineering Teams, ESETs. Teams of two to three members go to random Air Force sites and try to gain control of their computers. The teams have had a achievement rate of 30% in gaining complete control of the systems.

Indications

Commonly information is conveyed to incident response teams where there is not a specific victim or computer security incident but rather an indication that some activity may be ongoing in some other part of the society. There is presently no customary way of sharing that information, even though occasionally this information is posted to a shared list of FIRST members. This is one of the main areas that needs drastic improvement in the future in order to successfully route cyber crime and terrorist indicators that may be revealed within the incident response community.

New Technical Threats

The sharing of rising technical threats is one of the best successes of late. Today, the FIRST organization sponsors technical forums on an intermittent basis to discuss recent progress in technical threats and vulnerabilities. While this is an important meeting for the incident response community, it is not open to other international experts to relate the technical trends to political or other worldwide trends.

Difficulties in Tracking the Intruders

One noteworthy difficulty with the tracking of malicious activity to its origin is that it requires the cooperation of many autonomous administrative domains operating internationally. This makes the creation of a universally adequate technology solution unlikely (e.g., modification of the computer protocols and services to repeatedly identify the point of origin). Consequently a rational approach to track and trace today is the combination of technically experienced personnel working with these domains to gain their cooperation willingly to trace back malicious activity to the actual threat behind the activity.

With the size of information collected by the CERT/CC and other international response teams, it is likely that the network of these incident response teams will provide the best start to pursuing the activity of intruders and other malicious cyber movement to establish the point of origin. Although this can never be a definite activity, putting together all of the international data related to a malicious activity will provide the best cyber evidence of this point of origin and perhaps help to differentiate between domestic and international threats.

In addition to the information gathered by incident response teams, the CERT/CC has a distinctive affiliation with many of the Internet Service Providers (ISPs) and network operation centers that facilitate the CERT/CC to work with both victims and transport providers to track specific activity to its immediate point of origin. Then with its reputation of impartiality, the CERT/CC.

Now, as the world enters the Information Age, the nation's enemies may go after cyber targets. That is why the federal government is in the initial stages of evaluating the nation's telecommunications and information vulnerabilities, predicting that foes may look to strike Internet "network access points" instead of submarines and missile silos.

Government officials and members of the private sector meet for the first time Friday to start locating vulnerability points to potential cyber or physical attacks. The study, which may take years to complete, aims to recommend ways to eliminate those weak spots, create a system for identifying and preventing attacks, and prepare for attacks through training and education.

This study stems from a presidential directive, signed in May that created a Critical Infrastructure Assurance Office. The office will oversee the development of the first national plan to protect the services that the country depends on daily. But so many of those services – transportation, banking and finance, electric power, emergency services and so on – are becoming increasingly reliant on telecommunications and information components.

"It's kind of like a farmer who goes out after the winter storm looking to see where the fences are down and where the herd can get out," said Larry Irving, head of the U.S. Commerce Department's National Telecommunications and Information Administration, which is overseeing the telecom study. "It is kind of like surveying the situation. This is a new frontier for us. There is a new infrastructure out there. And we're much more dependent on this infrastructure than we've ever been or ever expected to be."

But Irving acknowledges that he needs the cooperation of private industry in order to conduct a thorough analysis of the infrastructure of a variety of industries. This study stems from a presidential directive, signed in May, which created a Critical Infrastructure Assurance Office. The office will oversee the development of the first national plan to protect the services that the country depends on daily. But so many of those services – transportation, banking and finance, electric power, emergency services

and so on – are becoming increasingly reliant on telecommunications and information components.

"While the meeting is open to the public, the results of the government study will not. In fact, in order to win the private sector's cooperation, government officials pledge that they will not keep information on file or divulge it to those who might exploit it.

"One of the things have to do is identify and remedy the situation," Irving said. "On the other hand, we want to make sure we're not putting information out there that could be used by terrorists or competitors. The last thing we want to do is create that road map."

The high-tech and telecommunications industries have recently been at odds with government on such issues as the export of encryption technology, and laws requiring telephone carriers to modify their equipment so law enforcement can carry out wiretaps over digital switches. But Irving said he is trying to foster the spirit of cooperation with industry in a comparable way to the work being done to remedy the year 2000 computer problem."⁹

In addition, the private sector may not have much choice. U.S. policy-makers in the White House, on Capitol Hill and in national security roles say the threat of information weapons – coming not only from terrorist operatives but also from foreign governments – is a potential danger. The targeting of infrastructure facilities with widely available cracking techniques could immobilize such network-connected services as electric power, banking and telephone. The exposure of both government systems and those in private industry has been underscored by the slew of attacks this year on everything from Pentagon computers to The New York Times' Web site.

In an article posted by Elizabeth Wasserman in "The Industry Standard", she observes¹⁰

"As the telecommunications structure is changing rapidly, Irving acknowledges that his agency's study will be a snapshot in time. But it will alert both government and private industry to the need for building protections into information and telecommunications systems. "We built an entire system of sidewalks with no curb carve-outs," he said. "If you're building the protections in as you go along, it's much easier."

The bottom line is that our world is ruled by information and by the people who manage it. In addition, no matter how correct the hackers think they are, the government

must protect itself and its citizens from future cyber attacks, whether they are harmless muscle flexing, or a full-fledged attack by another country.

Advantages and Disadvantages of Internet

The similar advantages the Internet and superior technology bring to the public and to business -- speed, security and global linkage -- are helping international terrorist groups arrange their lethal and disorderly activities.

Jay Lyman in his article "How Terrorists Use Internet", writes: ¹⁰

"The Internet and e-mail provide the perfect vehicles for these groups to communicate with each other, to spread their message, to raise money and to launch cyber attacks," defense director of intelligence for special projects Ben Venzke told NewsFactor Network.

A recent report from U.S. officials indicates that terrorists' use of the Web for communication and coordination through the use of encrypted messages is widespread, with numerous sites -- many of which are unaware of the use to which they are being put -- serving as conduits for terrorist conspiracies.

Government and private Internet security firms are doing their best to keep up with the terrorists, but the task is made more difficult by advancing technologies available to groups bent on targeting the U.S. and its citizens, allies and businesses'.

Officials in both government and private sector agree that terrorists such as Osama bin Laden and other extremist groups heavily use the Web, including Middle East terror organizations Hezbollah and Hamas.

"Terrorists use the Web mostly for propaganda and for information exchange," said Matthew Devost, founding director of the Terrorism Defense Center. "If you move beyond the Web, terrorist organizations do use information technology as a very viable and secure communication mechanism."

'Devost told NewsFactor that despite the Internet's viability as an economic medium, it has proven somewhat insecure for commercial transactions. He said the Web could help facilitate attacks by terrorist groups on not only the Internet economy, but on power, transportation and other systems that rely on information that is linked to the Web. Terrorists are beginning to use the Web in interesting ways, director of intelligence Jerry Freese told NewsFactor. "There's really no limit to it," Freese

said. "Anywhere you can send an e-mail with an audio or graphics file is fair game." Freese, whose security company provides secure servers, intruder detection and security audits, said terrorist cells around the world use the Internet for scheduling, meeting and organizing. "We see the Web as a terrorism-assistance tool that allows them to do things in secrecy," he said, referring to encrypted messages. "The thing is, it can originate from anywhere. The Web, of course, is ubiquitous."

'Freese said steganography -- putting encrypted messages in electronic files -- is widely used by terrorist groups. A recent government report indicated that terrorists have been hiding pictures and maps of targets in sports chat rooms, on pornographic bulletin boards and on Web sites. Despite their ongoing efforts to cripple parts of the Web, disrupt infrastructure systems such as electrical power or steal money and information from government and businesses, terrorists have a vested interest in keeping the Internet working. "It's a very good tool for them," Freese told NewsFactor, "so they don't want to disrupt the flow of the Web; rather, they" target specific companies that are working with or are sympathetic to their enemies."

'While law enforcement officials are aware of terrorists' use of the Internet, they cannot monitor Web sites for both logistical and legal reasons, according to spokesperson Steve Berry of the U.S. Federal Bureau of Investigations' National Infrastructure Protection Center. "However repugnant to our perception and to the general public and law enforcement their Web site or use of it might be, that does not give us the authority to block them," Berry told NewsFactor. "That's free speech. That's the country we live in." 'Venzke, whose company tracks Web-based threats for Fortune 500 companies and government, said law enforcement is also limited by national culture and geography. The Web offers entry into any country from anywhere, and with so many points linked together, terrorist activity is often impossible to track. "How do you force an [Internet service provider] halfway around the world, which may not be friendly to you to begin with, to shut down a Web site?" Venzke asked.'

The rapid advancement of technology makes it hard to fight terrorists, who, experts agree, are adept at using the Internet and other advanced technology. Bin Laden's al Qaida and other terrorist groups have reportedly used encryption programs available free on the Web, as well more powerful anti-spy software purchased on the open market.

The Terrorism Research Center's Devost said that despite a number of valid efforts to combat terrorists, targeted countries and businesses are not prepared.

"Most nations, and most companies, are not being diligent with regard to addressing information security concerns and fortifying their security posture," said Devost.

"Right now, it's very hard to detect where these messages are coming from and what their intent is," said Freese.

"Information exchange is a key issue here. We have many repositories of information, but it is not shared. The government is trying to collate information from private and government sources to coordinate defenses." Devost agrees, adding that despite increased efforts to keep tabs on terrorists, vulnerabilities are on the rise.

"Governments are making great progress in understanding the way these groups are utilizing technology," Devost said, "[but] while we are making progress, it is not enough."

Bob Sullivan in an article posted on MSNBC writes:¹²

"Envisage, for instance, if the Internet suddenly stopped working. A hacker group told Congress it could do it in half an hour. On the other hand, if power to major cities were interrupted. Government-hired hackers did that in four days in 1997. Alternatively, if parts of the 911 system were cut off. A Swedish hacker now in an asylum managed briefly to cut off 911 services in Florida two years back. Such "nuisance" hacks on communications are less spectacular than the hijacking of a missile, but they might be more valuable.

"The psychological impacts of IW (information warfare) can't be overstated," said Frank Cilluffo, director of the Information Warfare Task Force at the Center for Strategic and International Studies. "Using it, terrorist groups can achieve what they cannot militarily.

"We are the most technologically advanced country in the world, which means we have the most to lose," he added. "The United States is not very prepared to lose power, for example. And how long can you live without that database? What if suddenly all e-commerce were cut off?"

Putting banking into that e-commerce category; throughout the Kosovo clash, numerous reports indicated U.S. intelligence agencies had hired hackers to fiddle with international bank accounts full of Yugoslav President Slobodan Milosevic's money. There was a lot of debate in the security community about how likely this may be, but even the idea jolted the financial industry. Once that Pandora's

box is open — once one government's hackers are capable of freezing or changing personal bank account information — other governments and terrorist groups certainly would follow. Moreover, given that the entire banking system is established on confidence, such an attack could totally destabilize the reliability of the banking system, according to Kawika Daguio, executive vice president of the Financial Information Protection Association.

"Cilluffo's biggest concern is not an all-digital attack, but the use of computers as a multiplier for a more traditional attack. Imagine if a hacker had disabled 911 during the Oklahoma City bombing in 1995. Not only would medical help have been severely delayed, leading to more death and destruction — the resulting confusion would at least be demoralizing and, at worst, create a panic.

For proof of the potential for mob psychology, experts point to the Y2K bug. Even with several years' warning and continuous announcements that computers are Y2K-compliant, banks report cash hoarding has already begun, and survivalist-minded individuals are squirreling away water and dry goods.

"The actual problem is usually 10 times less damaging than the public perception of it," said Space Rogue, who runs the Hacker News Network service."

There is ample of debate about how intense the cyber threat is, despite the fact that recent signals from the U.S. government recommend federal agencies are taking it very critically.

"Just last week, The New York Times has leaked a document showing the National Security Council is working on a called the "Federal Intrusion Detection Network." The plan's director told the Times: "We know" foreign governments are developing cyber war capabilities, and "we have good reason to believe that terrorists may be developing similar capabilities." kill specific kinds of subjects.

"Not everyone is convinced the danger is very dramatic. After all, hackers did not gain access to the Pentagon's most secure systems. InfoWar.com founder Louis Cipher (a pseudonym) says Eligible Receiver and other high-profile cyber-threat incidents are part publicity stunt aimed at getting more federal money targeted to cyber warfare research.

"Paranoia is a bad thing, and America is being infected quickly," Cipher said. "Everybody's an alarmist.... You can disturb an infrastructure. Can go into telephony and can cause disturbance, a denial of service. But disturbing

electrical facilities is difficult. Just like on a railroad, they can go from track to track. There are a lot of safeguards.” And despite all the conjecture about cyber war capabilities, there is little evidence it has actually been used. In fact, even if the ability to take out power grids with a computer is out there, U.S. forces apparently showed a distinct reluctance to use the ability during the Kosovo conflict. So-called “soft bombs,” which short out electric lines, were used to create local power disruptions instead of a computer-based attack.

That satisfies Cilluffo, who thinks the United States should hold off crossing the line to cyber war for as long as possible.

“A well-placed bomb may still be easier,” Cilluffo said. “If we can go through physical means, then we are not compromising a technique that could be used against us.... After all, we have a lot more to lose.”

Conclusion

Cyber-terrorism is a complex issue that is vital for information security specialists and to some extent the society to recognize. One must be conscious of all of its characteristics in order to better evaluate how and where the "terrorists" are likely to attack our approach. What makes this subject significant is that a cyber terrorist attack is comparatively very easy and economical to instigate. For that reason, it can from any where in the world, at anytime, and, more importantly, the stakes of this cat and mouse game are can be quite high. In view of the fact that it is only a matter of time before a system is attacked, one must stay alert and share the knowledge obtained so that we can help protect ourselves at a personal, business, and national level.

References

- Crabtree, S. (1996) Cyberspace: a Terrorist Frontier? *Insight on the News*, 12 (31), 11.
- Cyber wars: Logic bombs may soon replace more conventional munitions. (1996). *The Economist*, 338 (7948), 77-78.
- DiDio, L. (1977). Feds Struggle in Race with Hackers. *Computer World*, 32 (15), 41-42.
- Laqueur, W. (1996) Terrorism: Looking to the Future. *Current*, 386, 9-14.
- Laquer, W. (1997) Terrorism via the Internet. *The Futurist*, 31 (2), 64-65.
- Pilant, L. (1997) Fighting Crime in Cyberspace. *The Police Chief* , 64 (8), 26-35.
- Smith, C. (1997) Get American Troops Ready to March on Cyber Battlefield. *Insight on the News*, 13 (44), 29-30.
- http://www.adl.org/internet/extremism_rw/before.asp The Consequences of Right-Wing Extremism on the Internet-Inspiring Extremist Crimes
- <http://www.seas.gwu.edu/student/reto/infowar/infowar.html> - An introduction to information warfare
- http://www.infowar.com/civil_de/civil_c.html-ssi - Detailed look at the dangers and future of cyber-terrorism.
- http://www.af.mil/news/feb1998/n19980206_980156.html - Analysis of hacking and cyber-terrorism.
- <http://www.terrorism.com/terrorism/index.html> - Main page of terrorism research center
- <http://www.kcl.ac.uk/orgs/icsa/> - Dr. Andrew Rathmell. "Cyber-Terrorism: The Shape of Future Conflict". *Royal United Service Institute (RUSI) Journal*. October 1997.

Notes

1. Cybercrimes: Infrastructure Threats from Cyberterrorists, *Cyberspace Lawyer*, 4 NO. 2 *Cyberspace Law*. 23
2. URL: <http://www.eeye.com/html/press/PR19990608.html>
3. URL:<http://www.fbi.gov/pressrel/pressrel101/nipc030801.htm>
4. URL:<http://www.cert.org/advisories/CA-2001-20.html>
5. Statement for the record by Michael A. Vatis, Director, National Infrastructure Protection Center (NIPC), Federal Bureau of Investigations (FBI), on NIPC Cyber Threat Assessment before the Senate Judiciary Committee, Subcommittee on Technology and Terrorism, October 6, 1999.
6. URL: http://www.adl.org/internet/extremism_rw/before.asp The Consequences of Right-Wing Extremism on the Internet-Inspiring Extremist Crimes
7. Author Unknown. "Cyber-terrorism" Document Date Unknown. URL:<http://www-cs.etsu-tn.edu/gotterbarn/stdntppr/index.htm>
8. Margaret Kane, "'Mafiaboy' busted in DoS attacks". *ZDNet News*.

9. Cmdr. Dave Pettinari. "Cyber terrorism, information warfare, and attacks being launched now and in the future in the heartland of America". Police Futurist International. Document Date Unknown. URL:<http://www.policefuturists.org/fall97/terror.html>
10. Elizabeth Wasserman. Feds take steps against threat of cyber terrorism. Industry Standard. September 25, 1998 Web posted at 12:05 PM EDT. URL: www.infowar.com
11. Jay Lyman. How Terrorists Use the Internet. NewsFactor Network. Posted September 12, 2001. URL: <http://www.newsfactor.com/perl/story/7731.html>
12. Bob Sullivan. Cyberwar: the threat of chaos. MSNBC. Document Date Unknown. URL: www.msnbc.com

The Impact of Computer Aided Design on Human Factors in Aviation Safety

Anne Curtney

Master of Aeronautical Science and candidate to PhD in Computing Technology at the Department of Engineering and Technology of the School of Doctoral Studies of the EU

[Email: a.curtney@det.sds.eu](mailto:a.curtney@det.sds.eu)

Abstract

This paper will review the impact of computer-aided design (CAD) upon aviation safety. In particular, the paper looks at how CAD has incorporated human factors principle - that is, the type of people who use the technology – into the design-making process. On the whole, CAD technology is an enormous improvement on older aviation design because it allows designers to visualize situations in three-dimensionality throughout the design process. CAD can provide a very good visualization overview of human interaction with the product (in this case, the pilot user interface of a cockpit).

Aircraft accidents that lead to loss of lives receive the highest level of attention in the aviation industry. There are many different causes for aircraft accidents that occur on an everyday basis in the aviation industry, which range from pilot errors to mechanical challenges. Some of those errors are minor, but others can have devastating effects. Pilot errors can be minimized by utilizing adequate tools and various analyses. Human factors are one of the methods that would improve safety. The term "human factors" has grown drastically, and has become more popular as the commercial aviation industry realized that human errors underlies most aviation accidents and incidents, rather than mechanical failure (Greaber 1999).

Human factors main focus is the safety of humans and their interaction with their surrounding work environment. Engineers are involved with a variety of disciplines; such as engineers, pilots, and mechanics in order to implement the latest advanced technology that would improve safety and efficiency in aviation.

Statement of the Problem

CAD technology is an excellent tool for designing pilot user interfaces because it saves time and money. Beyond that, CAD technology is infinitely superior to the research tools that preceded it, because it allows designers to see hypothetical human reactions to environmental stimuli in three dimensions. At the same time, CAD is not so intuitive that it can anticipate every conceivable human reaction that might result from a particular situation. Therefore, there remains a need for live testing with actual people.

The following study will investigate the impact of newly integrated CAD systems in improving safety factors in the aviation industry by looking at where it can be improved. Although designers now have a tool whereby they can design products to suit the physical abilities and/or characteristics of human operators, further research is needed to ensure that CAD realizes its full potential. This paper also will touch briefly on the CAD-based ergonomic analysis programs that allow researchers to create accurate 3-D human models that measure the impact of ergonomics on the safety and reliability of products, equipment, and facilities. This study hopes to underscore the manner and extent to which CAD technology can reduce product-related injuries and save money and lives, as well as the extent to which it remains dependent on human test subjects.

Limitations

Whenever a study examines new technology, there will be limitations. Today, CAD technology offers the promise of creating cockpit devices that will accommodate pilots of every level and idiosyncrasy. But while a CAD-based ergonomic analysis program allows for the creation of accurate, 3-D human models, it seems extremely unlikely that any computer program can truly simulate the abnormal responses of different human beings. Therefore, human subjects are desirable in assessing the effectiveness of different products within the cockpit – and finding an appropriate sample group is not easy. It is extremely difficult to assemble a representative sampling of different pilots (in simulated conditions, of course) with the intent

of measuring the capability of CAD in avoiding in-air tragedies.

The other potential problem is that when dealing with a product that accommodates and compensates for the physical abilities and characteristics of human operators, and if it is to attempt to determine the capability of that product, it necessarily stands to reason that a representative sampling of each kind or type of pilot is necessary. If there are pilots who suffer from certain medical conditions, for example, and determining the extent to which CAD technology permits them to escape harm only can be done if there is an adequate sampling of these types of pilots. General assumptions, however, can be made by studying a smaller sampling of pilots (in simulated conditions) who do not have identifiable medical conditions. This process is obviously narrow in scope, but it offers a tentative assessment of the capability of CAD technology.

Definition of Acronyms & Terms

CAD: the acronym CAD is short for computer-aided design. It involves the application of computer technology to the cockpit so that it is more ergonomic and more sensitive to the characteristics of different pilots.

CATIA: This was one of the first CAD programs that provided 3-D modeling of a human being. It seems improbable that CATIA can “ape” the precise reactions of a multitude of different human beings to environmental stimuli. CATIA, however, is certainly a promising tool for further research. The program runs on IBM mainframes, RS/6000 and HP 9000 workstations, and the expectation is that it will work with other platforms in the future (“CATIA definition,)

Human Factors Principles: This is an approach to technological design that incorporates the human element into the conception phase. It is not clear, however, whether the assumption is that human beings respond in the same ways to the same external stimuli, or if they or interact with program software in the same way.

NTSB: National Transportation and Safety Board (United States)

FAA: Federal Aviation Administration

TSA: Transportation Safety Administration

Review of Relevant Literature and Research

CAD technology does offer a great deal of promise for the future. Also can replace human subjects during the

testing process, being able to produce passable 3-D models of human subjects, CAD can reduce budgetary over-runs, save time, and save designers a great deal of stress, Springer et al (1991, p.375). One author notes that, “Because of the pressures placed on individuals in any high technology sector – and especially the commercial airline industry – this pressure extends to managers and supervisors” (Ungson, 1990, p 57).

Furthermore, because of the absorbing nature of piloting an aircraft, any process that can provide individualized tools for the individual pilot is a positive step. This endorsement, however, does not mean that just any kind of CAD technology is desirable: Addressing the effects of properly designed CAD technology on the human users of such technology, Backs and Boucsein (2000) write that “human performance and physiological costs depend on design and task-support characteristics of specific CAD technology.” (2000, p. 318). CAD technology – like any other innovation – only works if it is accompanied by skilled engineering and design. Still, Mark R. Rajai (2002) is one of a number of observers who believes that CAD technology can be adapted to produce “user-friendly” devices – as long as the incompatibilities between CAD and ergonomic programs are addressed, (2002).

Integrating Virtual Reality

Julian Berta is a thoughtful observer of the CAD scene. Assessing the technology’s relative merits and demerits, he offers a balanced assessment of the technology. Most conspicuously, he notes that CAD has traditionally offered only “non-intuitive, mathematically accurate” simulations, Berta (1999, p.14). The non-intuitive part is especially important, because it reveals a basic weakness of CAD’s that could prove deadly in the air: Non-intuitive technology is technology that, for the most part, is incapable of non-linear deductions vis-à-vis possible outcomes. Therefore, CAD technology is not overly effective – although this is changing, Berta (1999, p.14)– at anticipating unexpected outcomes. When one considers the possibility that the ergonomic design of a cockpit may fail to accommodate sudden, unexpected actions by a pilot, then the reliability of that technology to prevent a tragedy is greatly reduced.

Berta presents other shortcomings of the technology. For one thing, he notes that efforts to make CAD programs “immersive” have long run into difficulty because the process historically creates a loss of topological information as well as a loss of geometry or precision. More importantly, if certain “behaviors” have been assigned to objects during

CAD design, these behaviors are almost always lost during the conversion process, Berta (1999, pp.15-16). On the other hand, recent advances in CATIA suggest that efforts to make CAD more reliable are finally beginning to make progress, Berta (1999, p.18). And certainly, CATIA does allow researchers and designers to capture (most) interface and cockpit design flaws early in the creative process, Sarsfield et al (2000 p.50) – something that was largely impossible to do even a generation ago. On the other hand, there are still nettlesome questions. Gawron et al (2002) write that Safework, “which runs as a process under CATIA” is “a human modeling system that is strong in its application of geometrical data to mannequin size and shape” (p.191). Apparently, mannequins serve as the theoretical model for Safework – and Safework, as mentioned, often runs as a component of CATIA. The problem with using a mannequin as a model is that it is not a person and therefore only can approximate the physical features and characteristics of a person. While the matter may be trifling in the end, the complexity of the human organism would seem to make it dangerous to base computer models on a crude facsimile. Literature of this sort suggests that progress is being made, but it is ill advised to jump to the conclusion that the technology has now reached a stage where it is essentially infallible.

CADalyst

Joe Greco examines Human Factors Engineering. Greco's paper underscores how ergonomics is primarily the study of how human beings interact with machines (2004, p.46), but the strength of his paper lies in his detailed description of the many uses to which CAD technology can be put (pp.39-40). Greco does not privilege us with a review of how CAD technology can bolster aircraft safety, but one interesting observation he does make is that new algorithms now permit a closer “virtual reality” examination of “fuel-tank sloshing” and “underwater shock analysis” (2004, p.40). Greco does not explore these items in the depth one would hope, but a case can be made that innovations in computerized motion simulation can segue into a closer examination of how pilots might react in instances of severe turbulence and/or sudden impact. Given that new and more complex algorithms are being developed, the functional applicability of CAD technology is growing, as well. More time and research, however, is required before CAD technology can become all that its proponents hope for it.

To this point, the present literature has been fairly critical of CAD technology and critical also of any claim that CAD technology in its present form can serve as a bulwark against pilot error and avoidable mishaps. Yet while there is certainly room for skepticism, there is also much room for praise. Without question, aviation designers do benefit from CAD technology, and these benefits commonly involve reduced costs and time saved.

Visualization with CAD

As Ali E. Kashef (1991) notes, CAD simulations, whatever their imperfections, do offer designers convenience, added safety and what he calls “controllability” over real experience; additionally, computer-aided-technology systems allow for the revolution and rotation of objects within virtual space, and this definitely provides an element of three-dimensionality. Every bit as meaningfully, Kashef (1991 pp.64). Kashef writes that the application of CAD technology does not have to engender anxiety or a sense of dislocation among aviation engineers; the technology still requires “orthographic projection, descriptive geometry and other engineering graphic concepts” which have always constituted the foundation of the design process (1991, p.64). In other words, CAD has the potential to save costs by reducing (or even eliminating) the need for “live” test subjects – a view with which this writer does not entirely agree - but it does not necessitate a profound shift in how the design process is carried out.

Ergonomics and User Interfaces

CAD technology does bolster the ability of researchers to incorporate human dynamics into the design process, does allow for greater ergonomic design, and does allow for greater anticipation of the possible interactions of a pilot with his aircraft while in flight. In times of great stress, people will often forget that which they have labored years to learn. As a result, procedures which can serve as “memory aids” during these moments of crisis are invaluable to pilots of any caliber, Bullinger & Ziegler (1999, p.322). Equipment that is user-friendly and responds intuitively to the actions of a pilot in distress can save lives; at the very least, a pilot's user interface that has been well designed from a safety perspective can provide both pilot and co-pilot with far greater prospects for survival than might otherwise be possible. Yet Bullinger and Ziegler (1999) especially focus on advanced military user-face technology, which keeps

detailed data on a pilot's personal predilections and which then prompts the pilot accordingly – the fact remains that even advanced CAD technology demands “an empirical study (to) assess the effectiveness of the adaptations in an operational context”(1999,pp.684-685).In short, CAD technology is a wonderful graphical tool, but it is not yet so prescient that it can adequately anticipate all of the possible scenarios that might arise during the course of a flight.

Statement of the Hypothesis

It hypothesized that CAD technology is an excellent first step toward safer skies, but it must be complemented by the use of human subjects wherever possible. This study is informed by the general premise that CAD technology as applied to the aviation industry is quite literally a lifesaver. Moreover, the technology certainly can save money and time and those are valid issues when research labs operate under time constraints and bump up against the limits of finite budgets. This writer also believes that it is unwise to wholly cast aside human subjects during test simulations because computers are not so intuitive that they can possibly anticipate every action a pilot might make when confronted with a particular stimuli in flight. Bullinger and Ziegler have noted that military training takes into account the personal histories of pilots so that the pilot user-interface can be customized accordingly. While ergonomics in commercial airline companies has not reached that extent, it is certainly worth a closer look by concerned senior management, as even one mishap in the sky can be devastating.

Methodology

Research Design

The design for this project will be to assess the efficacy of CAD-based ergonomic analysis program through a comparative analysis in which a small sample population of pilots will be tested for their reaction to various simulated stimuli using pilot user-interfaces designed using CAD technology and Human Factors Principles. At the same time, CAD that employs only 3-D creations will be tested using the same pilot user-interface. The point of the exercise is this: If the pilots score lower than the 3-D projections (even though both the human subjects and the animated subjects are being tested for responses using the same computer technology), then the CAD technology is not doing an effective job of providing tools these pilots

can use in the air during times of crisis. Beyond that, if these findings are replicable across three or four tests for different stimuli, then it is clear that CAD technology has not advanced to such a degree that it can wholly do without the use of live human subjects. This being so, more data about human responses under certain conditions needs to be incorporated into the calculations of designers so that the interfaces that pilots use in flight simulators and in real life perform optimally.

Survey Model

The survey will be separated into two parts. The first part will be simulated tests in which 3-D models created by CAD technology and CATIA programs will respond to computer-designed hypothetical situations; these 3-D models will use the same user interfaces their human counterparts will be using. Afterward, for the second part of the research study, there will be a sampling of human subjects who will use the same interfaces and will be asked to respond to the same stimuli. Because an actual pilot only receives one chance to correctly respond to stimuli in real life, there will be only one test administered to the human participants for each stimulus. The scores between the two groups will be collected and averaged and standard deviations calculated for the respective 3-D and human results; if the human scores are lower than the 3-D animation scores, then the interface or cockpit design is not working as well as it might for actual human pilots.

Survey Population

As noted above, this paper would ideally prefer to have commercial pilots involved in the testing, but that could prove costly and/or impracticable. At the same time, trained pilots are vital to the exercise because they are the ones for whom the cockpit interfaces have been designed. Therefore, every effort will be made to secure the services of a small sample population of pilots who can participate. If none are available, then the project will be modified and it will test only cockpit interfaces designed for novice flyers. At this point, the human sample group consists of novice pilots who have volunteered their time. A large sample group would be preferable so that there is no possibility of the findings falling outside the statistical margin of error, but it is anticipated that only two to three dozen people will be enlisted, given the practical difficulties involved in securing people's time.

Results

All told, fifteen (15) novice flyers agreed to participate in the study in time for inclusion in the statistical analysis. A comparison of the results using 3-D simulated tests and human subjects is provided below and discussed further in Chapter VI below.

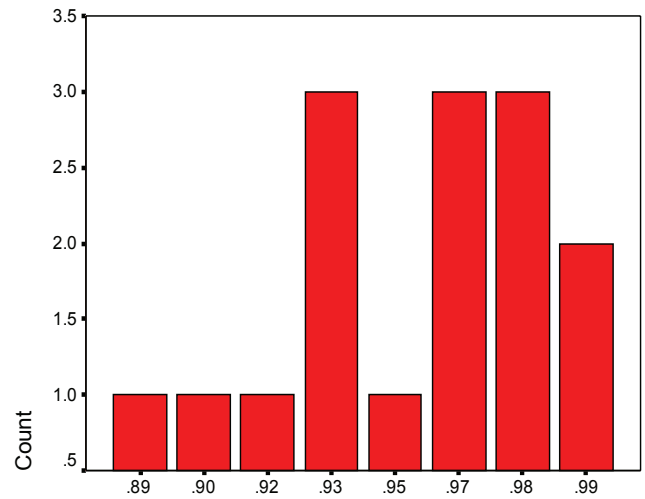
Case Summaries

Simulated tests: 3-D models CAD/CATIA programs
Simulated tests: Human novice pilots

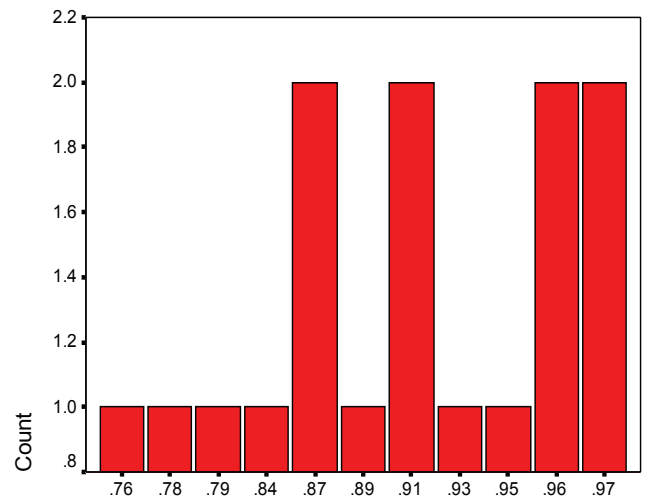
1	.98	.87
2	.93	.93
3	.89	.91
4	.92	.96
5	.97	.95
6	.98	.84
7	.99	.79
8	.98	.87
9	.97	.97
10	.93	.91
11	.93	.89
12	.95	.76
13	.97	.96
14	.90	.78
15	.99	.97
Total N	15	15

Simulated tests: 3-D models CAD/CATIA programs

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
.89	1	6.7	6.7	6.7
.90	1	6.7	6.7	13.3
.92	1	6.7	6.7	20.0
.93	3	20.0	20.0	40.0
.95	1	6.7	6.7	46.7
.97	3	20.0	20.0	66.7
.98	3	20.0	20.0	86.7
.99	2	13.3	13.3	100.0
Total	15	100.0	100.0	



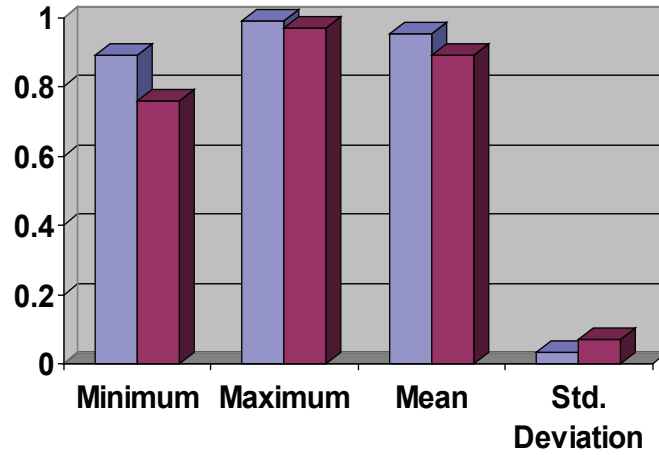
Simulated tests: 3-D models CAD/CATIA programs



Simulated tests: Human novice pilots

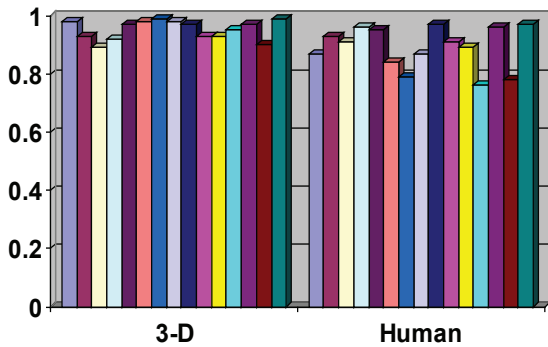
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Simulated tests: 3-D models CAD/CATIA programs	15	.89	.99	.9520	.3299
Simulated tests: Human novice pilots	15	.76	.97	.8907	.07116
Valid N (listwise)	15				

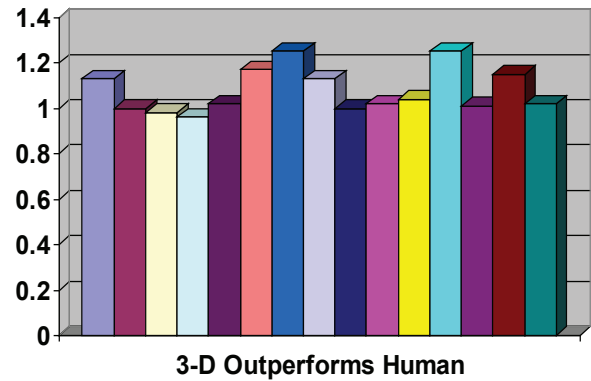


■ Simulated tests: 3-D models CAD/CATIA programs ■ Simulated tests: Human novice pilots

Test #	Simulated tests: 3-D models CAD/CATIA programs	Simulated tests: Human novice pilots	3-D vs. Human
1	0.98	0.87	1.13
2	0.93	0.93	1.00
3	0.89	0.91	0.98
4	0.92	0.96	0.96
5	0.97	0.95	1.02
6	0.98	0.84	1.17
7	0.99	0.79	1.25
8	0.98	0.87	1.13
9	0.97	0.97	1.00
10	0.93	0.91	1.02
11	0.93	0.89	1.04
12	0.95	0.76	1.25
13	0.97	0.96	1.01
14	0.9	0.78	1.15
15	0.99	0.97	1.02
Total	15	15	
Average Score	0.95	0.89	1.08



■ 1 ■ 2 ■ 3 ■ 4 ■ 5 ■ 6 ■ 7 ■ 8 ■ 9
 ■ 10 ■ 11 ■ 12 ■ 13 ■ 14 ■ 15



■ 1 ■ 2 ■ 3 ■ 4 ■ 5 ■ 6 ■ 7 ■ 8 ■ 9 ■ 10 ■ 11 ■ 12 ■ 13 ■ 14 ■ 15

Discussion

A number of changes have taken place in design of aircraft in general and cockpits in particular in recent years that will likely have a profound effect on the manner in which these systems are tested and evaluated for appropriateness and effectiveness as to human usability in the future. Today, in spite of miniaturization and digital displays, there is still just so much space available in the cockpit in which to fit all of the instruments required to keep a pilot situationally aware. Until the 1960s, the vast range of telemetry required for the safe operation of aircraft available to the human pilot typically assumed the form of a single-purpose instrument as shown in Figure ____ below, which illustrates a cockpit from this time period.

Figure _____. Typical cockpit design of the early 1960s, as illustrated by the X-15.

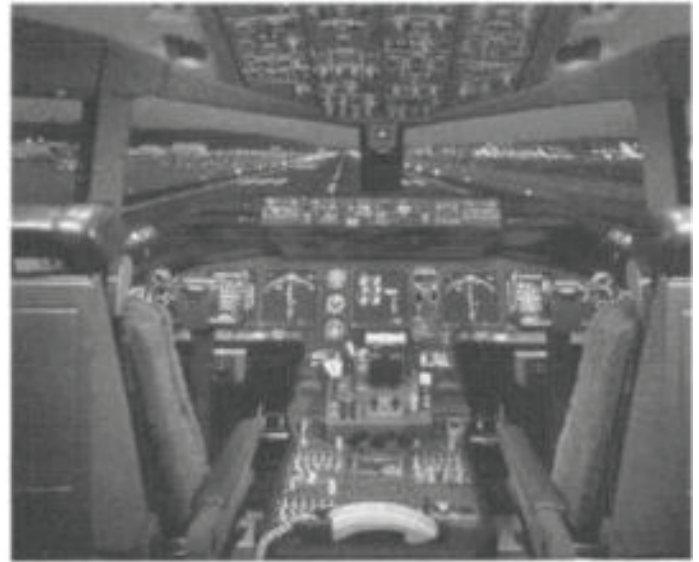


Source: Tsang & Vidulich, 2003, p. 6.

This elaborate array of indicators tended to exert a serious visual scanning and cognitive integration load on the pilot; furthermore, as aircraft became increasingly complex and the amount of information that was needed to be presented to the pilot increased, and the finite "real estate" of the cockpit became a seriously constraining design factor. As these authors put it, "A point in evolution was reached in which the number of pointers, numeral counters and flags could no longer be increased or their display characteristics improved" (Tsang & Vidulich, 2003, p. 6). In sharp contrast to the 1960s cockpit version, the prominent role of multifunction displays in modern aircraft

cockpits is highly evident as illustrated by the Boeing 777 cockpit shown in Figure ____ below.

Figure _____. The Boeing 777-300 flight deck.



Source: Tsang & Vidulich, 2003, p. 6.

According to Tsang and Vidulich (2003), the transformation from single instruments and gauges to more flexible multifunction displays has resulted in a dramatically different appearance for modern cockpits compared to their 1960s counterparts; however, a great deal more has also changed besides just the way the cockpit looks: "In addition to the changes in display technology, the 1960s ushered in the use of computerized automation to assist pilots. For example, the Boeing 777 computer systems incorporate more than 2.6 million lines of software code to support the autopilot, flight management, navigation, and maintenance functions" (Tsang & Vidulich, 2003, p. 7).

Furthermore, the introduction of highly automated aircraft featuring so-called "glass cockpits" in recent years has also resulted in some mixed blessings. On the one hand, these innovations have extended the capabilities of the aircraft but on the other hand, they have also transformed the nature and type of tasks that have to be performed by the pilot. Today, flight control assistance and flight management systems (FMS) have changed the pilot's role from that of a manual controller and navigator to one of systems monitor and information manager in a highly dynamic environment (Amalberti & Sarter, 2000). According to these authors, "Information and resource management, task scheduling, and programming skills of onboard computers now complement the psychomotor

skills of the pilot. The continuing expansion of air travel necessitates the use of even more advanced technologies in order to accommodate the expected levels of traffic while at the same time maintaining safety. Technologies like the digital datalink play a key role in realizing such advancement” (Amalberti & Sarter, 2000, p. 181).

Today, new, or retrofitted, cockpit equipment and human-machine interfaces must be applied in a highly complex operational context that is characterized by a so-called theoretical "timeline." Such timelines represent a mission profile specifying when, by whom, and in which order tasks should be performed. During the design process, timelines are used to assist cockpit design and define requirements for the input and display devices; however, such analyses are strictly theoretical in nature and are constrained in reality by variations in equipment use and extended response times (Amalberti & Sarter, 2000).

Other innovations have emerged in recent years as well that will undoubtedly have an impact on how the cockpit of the future is designed and how it is tested and evaluated. According to Charlton and O’Brien (2002), “Computer-based manikins or digital human models are three-dimensional, computer-graphic representations of the human form. Human modeling systems are interactive software simulation tools that combine three-dimensional human geometry with three-dimensional product geometry, obtained from a CAD system for the purpose of evaluating how well a particular design will accommodate the people who will be interacting with it” (p. 188).

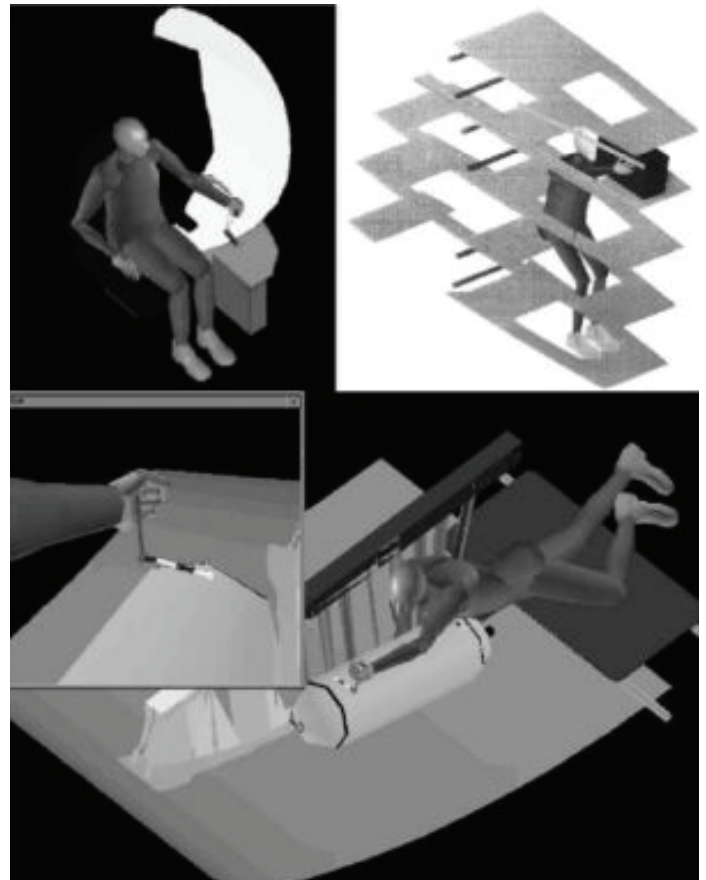
Computer-based manikins are controlled by human modeling systems to provide a variety of functions, including the ability:

1. To import and export product data/geometry;
2. To simulate human functions, such as reach and vision;
3. To perform analyses, such as access and interference analyses; and,
4. To provide support functions, such as reports, documentation, and user help (Charlton & O’Brien, 2002).

Figure ___ below illustrates how a reach envelope is typically used by cockpit designers to evaluate control placement. In this case, the reach envelope provides a quantitative description of the pilot’s maximum reach, assuming certain seat and restraint needs required by the task involved (Charlton & O’Brien, 2002). According to these authors, “The designer then uses this quantitative description to define the requirement that all critical controls be placed within the reach envelope to ensure easy access

and operation. Verification of a design can be accomplished by merging the reach envelope with a candidate design to ensure controls are within the envelope” (Charlton & O’Brien, 2002, p. 189). As also shown in Figure ___ below, a comparable procedure can also be used to assess visual and physical access (Charlton & O’Brien, 2002).

Figure ___. Using a computer-based manikin to assess a reach envelope (top left), physical access (top right), and visual access through the manikin’s eyes (bottom).



Source: Charlton & O’Brien, 2002, p. 189.

No matter how sophisticated and refined three-dimensional and CAD-assisted applications become in the future, there is going to be a fundamental need to ensure that the human element is not lost in the equation. According to Hanson (2001), “A key factor in the design of modern aircraft cockpits is ensuring that the target population can be fully accommodated. The cost of training, the scarcity of potential pilots, and the drive for maximising the potential for export have led to the expansion of the aircrew anthropometric population at both ends of the distribution” (p. 85). This author adds that the introduction of female

pilots has complicated issues for cockpit designers because they are typically smaller pilots with reduced strength capabilities compared to their male counterparts, and may not be able to manipulate the controls in the fashion expected by a computer-generated model. In this regard, these computer-modeled evaluations would be extremely limited in scope and breadth if they failed to take into account such gender-related factors in cockpit design, as well as how different genders (and potentially cultures) perceived the design elements of a cockpit as to functionality and effectiveness for their respective purposes. For example, as Hanson (2001) emphasizes:

It is of prime importance that all pilots are able to attain displacement of the stick throughout its full range of movement whilst maintaining necessary precision. Failure to do so could compromise the aircraft mission and endanger the pilot. However, while stick forces must be low enough for the smaller and weaker pilots to handle, they must be high enough for the stronger aircrew to maintain a good feel of the aircraft, without the risk of over-controlling. The issue of strength in cockpit design is an important one, and one which has been largely overlooked. (p. 85)

Therefore, while a three-dimensional model of a pilot might determine that everything is appropriate and efficiently designed for a computer-generated or manikin model, a living, breathing human being might find differently is ways that cannot be foreseen or preprogrammed into such evaluation applications. Although the findings of the experiment using computer-based evaluations versus novice human pilots supports the continued use of these applications in the future, there were some isolated instances in which the human subjects outperformed the computer models and these issues are discussed further below.

Findings from 3-D Model versus Human Pilot Experiment. The results of the experiment showed that the most common scores for both the computer model and human subjects were .93, .97, and .98 with three each (or 20 percent of the total 30 tests). There were two instances of a .99 score (or 13.3 percent) with the remainder of the scores of .89, .90, .92, and .95 representing unique scores (or 6.7 percent each) by the human or 3-D model.

The minimum score for the 3-D model was shown to be .89 compared to the human minimum score of .76, a maximum score of .99 for the 3-D model compared to .97 for the human subjects, and an average score of .9520 compared to .8907 for the humans. The 3-D model scores had a standard deviation of .03299 compared to 0.7116 for the human pilots.

The average score for the 3-D model was 0.95 compared to 0.89 for the human subjects, with the computer model outperforming the novice pilots by 10.8 percent across the board (human subjects outperformed the computer model in two instances).

Conclusion

The research showed that the 3-D visualization applications and CATIA software all provide significant design features that will undoubtedly prove valuable to cockpit designers in the future by making the process more efficient and responsive to human needs. The research also showed that in order to achieve these goals, it is vitally important to ensure that human subjects are part and parcel of the testing regimen, with even the best programs being limited to a finite number of variables that cannot be duplicated with any degree of accuracy in a computer model today. Just as weather may not ever be able to be accurately predicted using computer models because of the enormous number of variables involved (with weather representing just one of the many confounding factors that may adversely affect a pilot's ability to fly safely), computer modeling of all of the potential events that may occur during the flight management process is also currently not feasible. While many proponents of unmanned space flights to other planets because of the logistical nightmares involved in putting humans in space, the fact remains that real live people will likely pilot the first substantive missions to Mars and beyond because humans are required to respond to the unforeseeable that cannot be modeled in virtual terms. Likewise, when humans take to the skies, the cockpit in which they operate must be designed with their unique needs in mind, rather than what a computer model suggests is the most efficient solution.

Recommendations

Pilots of any type of aircraft must gather a wide range of telemetry that is used to develop a sense of situational awareness which is achieved through the use of various flight instruments and the navigational equipment on board; however, the process of constructing an accurate mental model of the position of the aircraft in space, and its condition is subject to a number of degrading factors that can adversely affect a pilot's ability to fly safely such as distraction, inattention, under-arousal, stress, boredom and fatigue. In these dynamic and potentially threatening environments, ensuring the accuracy of the telemetry

being received concerning the flight assumes increasing importance for its safe and effective management. According to one authority, "The abstract nature of mental models and situational awareness make them generally difficult to observe. . . . Fatigue effects demonstrate the importance of investigating mechanisms rather than tasks. In general, task-induced fatigue effects tend to be more detrimental when the task is attentionally demanding" (Desmond & Hancock, 2001, p. 21).

By ensuring that human subjects are incorporated into the analytical process to the maximum extent possible, cockpit designers can take advantage of the unique insights that only humans can provide to ensure that as many variables as possible are taken into account during the cockpit design phase. Studies have shown time and again that investments in the design phase pay major dividends in terms of improved flight management, and it is unreasonable to suggest that there may be a trade-off available in terms of saving human lives. Therefore, it is recommended that policymakers at all levels seek to ensure that timely and relevant guidelines are available for aircraft designers that require the use of human subjects in any computer-based application.

References

- Adler, P.S. (1992). *Technology and the future of work*. New York: Oxford University Press.
- Charlton, S. G., & O'Brien, T. G. (2002). *Handbook of human factors testing and evaluation*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cognitive Engineering in the Aviation Domain. Contributors: Rané Amalberti - editor, Nadine B. Sarter - editor. Publisher: Lawrence Erlbaum Associates. Place of Publication: Mahwah, NJ. Publication Year: 2000
- Backs, R. W. & Boucsein, W. (Eds.). (2000). *Engineering Psychophysiology: Issues and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Berta, J.(1999, September/October). Integrating VR and CAD. *IEEE Computer Graphics and Applications*, 19(5), pp.14-19.
- Bullinger, Hans-Jorg & Jurgen Ziegler. (1999). *Human-Computer Interaction: Ergonomics and User Interfaces (Vol.I)*. Mahwah, NJ: Lawrence Erlbaum Associates.
- CATIA definition. (n.d.). The Free Dictionary by Farlex. <http://computing-dictionary.thefreedictionary.com>
- Charlton, S. G., & O'Brien, T. G. (2002). *Handbook of human factors testing and evaluation*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Desmond, P. A., & Hancock, P. A. (2001). *Stress, workload and fatigue*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Federal Aviation Administration. (2003). *Administrator's Fact Book*. United States Department of Transportation: Federal Aviation Administration.
- Gawron, V. J., Dennison, T. W., & Biferno, M. A. (2002). Chapter 9 Mock-Ups, Models, Simulations, and Embedded Testing. In *Handbook of Human Factors Testing and Evaluation*, Charlton, S. G. & O'Brien, T. G. (Eds.) (pp. 181-218). Mahwah, NJ: Lawrence Erlbaum Associates.
- Graeber, C (1999). The Role of Human Factors in Improving Aviation Safety. Boeing Commercial airplanes Group .Aero magazine no. 8
- Greco, Joe. (2004). Human factors engineering. *CADalyst*, 21(7), pp.38-46.
- Hanson, M. A. (2001). *Contemporary ergonomics 2001*. London: Taylor & Francis.
- Johnson, Alan. (2005). Still some reluctance with CAD technologies. *Manufacturers' Monthly*, p.22.
- Kashef, A. E. (1991). Visualization with CAD. *Technological Horizons in Education Journal*, 19 (5), pp.64.
- Koonce, J.M. (1999). A historical overview of human factors in aviation. In D.J. Garland, J.A. Wise & V.D. Hopkin (Eds.), *Handbook of Aviation Human Factors* (pp.3-15). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rajai, M R. (2002). Development of an integrated ergonomic CAD system for evaluating designs for human factors and safety. LookSmart Education.
- Sarsfield, L. P., Stanley, W. L., Lebow, C. C., Ettetdgui, E., & Henning, G. (2000). *Safety in the Skies: Personnel and Parties in the NTSB Aviation Accident Investigations: Master Volume*. Santa Monica, CA: Rand.
- Springer, J., Langner, T., Luczak, H., & Beitz, W. (1991). Experimental Comparison of Cad Systems by Stressor Variables. *International Journal of Human-Computer Interaction*, 3(4), 375-405.
- Stanney, K M. *Handbook of virtual environments: design, implementation and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tsang, P. S., & Vidulich, M. A. (2003). Principles and practice of aviation psychology. Mahwah, NJ: Lawrence Erlbaum Associates.
 Ungson, G. (1990). International competition in high technology: the case of the United States, Japan and Europe In Susan Albers Mohrman and Mary Ann Von Glinow (Eds.), Managing Complexity in High-

Technology Organizations. New York: Oxford University Press.
 U.S. Congress Office of Technology Assessment, & United States. Congress. Office of Technology Assessment. (1984). Computerized Manufacturing Automation: Employment, Education, and the Workplace. Washington, DC: U.S. G.P.O.

Appendix I

SPSS Data Output File/Summarize

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Simulated test: 3-D models CAD/CATIA programs	15	100.0%	0	.0%	15	100.0%
Simulated tests: Human novice pilots	15	100.0%	0	.0%	15	100.0%

a. Limited to first 100 cases.

Case Summaries

	Simulated tests: 3-D models CAD/CATIA programs	Simulated tests: Human novice pilots
1	.98	.87
2	.93	.93
3	.89	.91
4	.92	.96
5	.97	.95
6	.98	.84
7	.99	.79
8	.98	.87
9	.97	.97
10	.93	.91
11	.93	.89
12	.95	.76
13	.97	.96
14	.90	.78
15	.99	.97
Total N	15	15

a. Limited to first 100 cases.

Frequencies

Statistics

		Simulated tests: 3-D models CAD/CATIA programs	Simulated tests: Human novice pilots
N	Valid	15	15
	Missing	0	0

Frequency Table

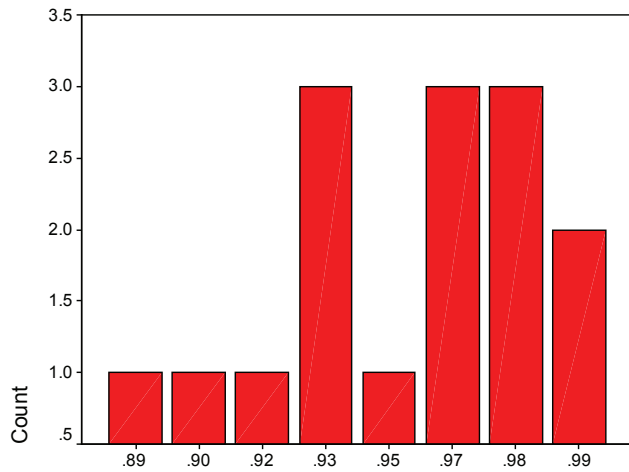
Simulated tests: 3-D models CAD/CATIA programs

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.89	1	6.7	6.7
	.90	1	6.7	13.3
	.92	1	6.7	20.0
	.93	3	20.0	40.0
	.95	1	6.7	46.7
	.97	3	20.0	66.7
	.98	3	20.0	86.7
	.99	2	13.3	100.0
Total		15	100.0	

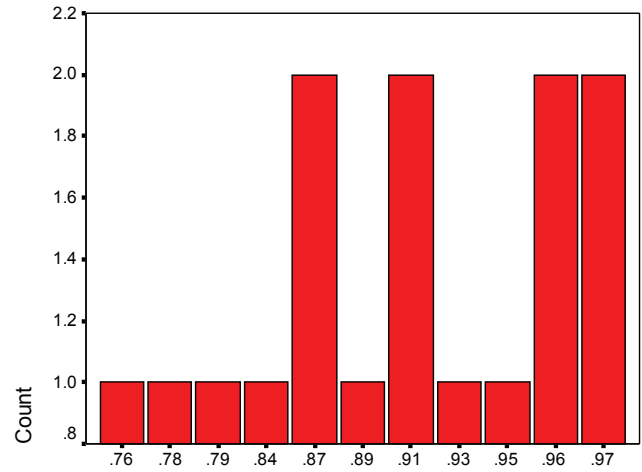
Simulated tests: Human

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.89	1	6.7	6.7
	.90	1	6.7	13.3
	.92	1	6.7	20.0
	.93	3	20.0	40.0
	.95	1	6.7	46.7
	.97	3	20.0	66.7
	.98	3	20.0	86.7
	.99	2	13.3	100.0
Total		15	100.0	

Graph



Simulated tests: 3-D models CAD/CATIA programs



Simulated tests: Human novice pilots

Appendix II

Excel Spreadsheet Analysis

	Simulated tests: 3-D models CAD/CATIA programs	Simulated tests: Human novice pilots	3-D vs. Human
1	0.98	0.87	1.13
2	0.93	0.93	1.00
3	0.89	0.91	0.98
4	0.92	0.96	0.96
5	0.97	0.95	1.02
6	0.98	0.84	1.17
7	0.99	0.79	1.25
8	0.98	0.87	1.13
9	0.97	0.97	1.00
10	0.93	0.91	1.02
11	0.93	0.89	1.04
12	0.95	0.76	1.25
13	0.97	0.96	1.01
14	0.9	0.78	1.15
15	0.99	0.97	1.02
Total	15	15	
Average Score	0.95	0.89	1.08

Section DSc

Science

Section coordinated by the Department of Science in the School of Doctoral Studies of the European Union

Head of the Department:

Professor Mark T. Abel

Reviewers Body Coordinators:

Chair of Mathematics:

Professor Mark W. Boulat

Deputy Head of Department – Earth and Biology:

Professor Alexandra Moffett

Chair of Earth and Environment Science:

Professor Sergio Farlow

Chair of Biology and Life Science:

Professor Melissa Brant

Chair of Physics and Astronomy:

Professor Timothy Olson

Chair of Chemistry:

Professor Randolph Laman

DSc Section Content

Article 1

Page 112

Enhanced Wellbeing amongst Engineering Students through Nadi Shodhan Pranayama (Alternate Nostril Breathing) Training : An Analysis

Author: Anurag Joshi (M.E.), Mandeep Singh (Ph.D.), Bharat Bhushan Singla (Ph.D.), Sunil Joshi (M.D.)

Country: India

Article 2

Page 121

An Effective Neuroprotective Treatment in Ischemic Stroke and Cerebral Trauma with Low Doses of L-Arginine, Lamotrigine and Tianeptine

Author: Héctor Jara (MD), Francisco García (MD), Franky Torres (MD), Fuad Lechin (MD,PhD)

Country: Venezuela

Article 3

Page 134

Metastatic Breast Cancer: An evaluation of the Safety and Effectiveness of High Dose Chemotherapy with Autologous Stem Cell Transplant

Author: Martin F. Taylor (MSc)

Country: UK

Article 4

Page 144

Natural Gas Sweetening Process Design

Author: Kurt Mearkeltor (MSc)

Country: Sweden

Enhanced Wellbeing Amongst Engineering Students Through Nadi Shodhan Pranayama (Alternate Nostril Breathing) Training : An Analysis

Anurag Joshi

M.E. (Electrical Engineering) Department of Electrical Engineering,
Thapar Polytechnic College, Patiala- 147004 (Punjab) India

Dr. Mandeep Singh

Ph.D. (Telecardiology) Department of Electrical and Instrumentation Engineering,
Thapar University, Patiala-147004 (Punjab) India

Dr. Bharat Bhushan Singla

Ph.D. (Management) School of Management Studies,
Punjabi University, Patiala-147002 (Punjab) India

Dr. Sunil Joshi

M.D. (Ayurveda Medicine) Babe Ke Ayurvedic Medical College & Hospital,
VPO Daudhar, Via Ajitwal, Distt. Moga (Punjab) India

Abstract

The state of wellbeing is determined by individual's Physical, Mental and Emotional health. This paper introduces the concept of pranayama (an ancient yogic technique that involves controlled deep rhythmic breathing). The underlying principle of pranayama is that the relation between emotion and breathing is two way i.e. not only emotions have affect on breathing but controlled rhythmic breathing has positive effects on emotions too. We selected a group of engineering students who volunteered to practise alternate nostril breathing i.e. Nadi Shodhan Pranayama (NSP) for three months. We applied the introspection (subjective observation) method of Psychology and analyzed the various traits related to wellbeing of the group on Likert's five point psychometric scale before and after applying this technique. We applied t-test for statistical investigation. We observed that 75% of the subjects gained in terms of Feeling Healthy, 80% in terms of memory recall, 75% in terms of mental stress relief and 90% in terms of physical relaxation. This amazingly simple and yet highly effective ancient technique of NSP may become part of their physical training routine to be followed regularly. The paper aims at spreading the awareness of this yogic technique on the wellbeing of all human beings in general and engineering students in particular.

Key words: Improved Wellbeing; Alternate Nostril Breathing; Nadi Shodhan Pranayama, Students.

Improved wellbeing is the combination of feeling good (including positive emotions such as happiness, contentment, interest and affection) and functioning well. We wanted to know the discernment of the students about wellbeing. We conducted a survey amongst the students, their parents and their teachers too to know their perspicacity about wellbeing. The students view was considered primarily where as the view of the parents and teachers was to endorse the view of students. The major aspects of improved wellbeing reported by the students after the survey were Feeling Healthy, Memory Recall,

Mental Stress Relief and Physical Relaxation. Feeling Healthy stands for freedom from corporeal disorder. If a student suffers any physical anarchy, he may neglect the study efforts and may not perform well in the examination. Memory power plays most important role for getting good grades in the examination. Hence achieving success in the examination may be indicative of student's wellbeing. Mental Stress is caused due to examination phobia and peer pressure as well, when they have to compete for anything. Also infatuation in adolescents and fear of rejection may play a vital role to enhance Mental Stress. Lack of physical

relaxation determines the performance in the workshop jobs or playground activities due to which they tend to experience fatigue and want quick relief from that. Also survey revealed that some students may hook on to drugs to attain state of wellbeing temporarily as they are not mature enough to differentiate between right or wrong. This may be due to their incorrect thinking and unawareness.

On basis of the survey reporting we decided the four parameters i.e. Feeling Healthy, Memory Recall, Mental Stress Relief and Physical Relaxation to determine the state of wellbeing of engineering students.

Technique

Ancient yogic technique of alternate nostril breathing also known as Nadi Shodhan Pranayama (NSP) is able to create the feeling of being well and this aspect of NSP is explored analytically in this paper. How the subject feels well is a matter of self introspection and cannot be measured by any instrument as such. One of the ways to know the wellness feeling of the subject is to ask to report the state of mind based on selected subjective parameters such as Feeling Healthy, Memory Recall, Mental Stress Relief and Physical Relaxation before and after NSP training. It may look that the last parameter physical relaxation is a physical state and not a mental state but on deeper instructions it is found that a person may be physically tired but may report as physically relaxed when he is emotionally positive. For example if a person is about to receive a gold medal in front of his colleagues, he may report physically relaxed even though he may have travelled overnight without any sleep. However to the same person if he is attending a close relative in Intensive Care Unit of the Hospital may report physically worn out even if he had all the physical features of rest and comforts of air conditioned environment. It is interesting to note that all diagnostic tests are recommended by the doctor only when the subject first feels diseased in his being and consults the doctor. Based on these simple facts the authors decided to make the subject self introspection report on the above cited four parameters before and after NSP training and analysed the results to know the efficacy of this ancient technique of feeling well.

Yoga

Yoga has a rich base to deal with physical as well as mental health. This is perhaps the reason that our Sages

used to lead a healthy and cheerful long life. If one practises yoga, one can surely notice how certain asanas affect one's mind and body. Some poses may be easy to attain and put our mind at ease, while other yoga poses are difficult and may even cause us frustration. These physical and mental feelings also help us in choosing the type of yoga that best suits us. Health disorders related to Blood Pressure, Heart, Lungs, Kidneys, Joints etc. are on the rise throughout the world [1]. It is well established fact that physical health is influenced by mental wellbeing and vice versa. Mental health is primarily more affected by negative emotions like lust, anger, infatuation, greed, pride, anxiety, fear, depression and inferiority complex [2]. Learning the management of these negative emotions is therefore essential to maintain good mental health. Yoga offers comprehensive solution for managing the negative mental emotions. It is an ancient Indian term signifying union of an individual with cosmic being. This is achieved by following various techniques in a disciplined way. The focus of this paper is on Pranayama (Controlled Yogic breathing). The underlying principle of Pranayama is that the relation between emotion and breathing is two ways i.e. not only emotions have effect on breathing but controlled rhythmic breathing have positive effects on emotions too.

Pranayama

Breathing helps in maintaining the vital energy of life and thus in yogic terms this is known as Prana. The process of controlling the Prana is called Pranayama. So pranayama is the science related to invigorating the vital force supplying energy and controlling the mind-body complex. The ancient texts emphasise that retention of air, increases the level of prana (energy) in the body, it also regulates the flow of pranic energy through out the body. So pranayama helps to control most of the ailments and can also slow down the inevitable aging process of the body [3].

The mind influences most of the endocrine and other physical as well as the metabolic functions of the body including breathing. When the mind is calm and relaxed, the breathing is smooth and slow. If one is stressed breathing is fast or shallow. In this way, the mental and emotional state of a person has a positive or negative impact on health through breathing [3].

The functioning of all the organs like heart, brain, digestive organs, endocrine glands in the body have certain rhythms. Similarly the breathing also has a rhythm.

Pranayama is deep rhythmic breathing bringing the breath in desired rhythm by controlling the process of inhalation, retention and exhalation [3].

In the process of breathing, the diaphragm, intercostal muscles and accessory muscles of respiration are used. The diaphragmatic breathing is called vertical breathing and is considered to be more efficient way to inhale air than inhaling while expanding the chest which is considered to be the horizontal breathing which involves simply expanding the chest. In pranayama, one should utilize the diaphragm efficiently to get more oxygen without making much effort. The diaphragm is attached to the organs like heart and lungs on superior surface and to the liver, spleen, pancreas and stomach on inferior surface. Efficient movement of the diaphragm makes the functioning of these organs more efficient [3].

Many researchers and Yogis have reported the benefits of practising pranayama on Diabetes Mellitus [4], Heart Rate [5] and Nervous System [6]. Also research through Yoga Meditation has shown remarkable improvement in Patience, Physical Relaxation, Mental Stress relief [7] and physical relaxation [8] of the chosen subjects.

Nadi Shodhana Pranayama (Yogic breathing control) has beneficial effects on Autonomic Nervous System. It decreases sympathetic discharge, lowers metabolic rate and increases parasympathetic discharge. This study also proves long term beneficial effects to body on stressors if yogic breathing exercises are practiced regularly. [9]

This systematic breathing pattern results in improvement in Feeling Healthy, enhancing Memory Recall, decreasing Mental Stress and imparting Physical Relaxation.

Methodology

Pranayama techniques are best practised while sitting on the floor on a blanket or a carpet or a mat. This form of practice is applicable to padmasana also. Padmasana is a posture in which the subject sits in a cross legged position as shown in the figure P1. However, any other posture is also acceptable provided the back is kept erect from the base of the spine to the neck and perpendicular to the floor. Bad and poorly performed posture leads to shallow breathing and low endurance. One must empty the bladder and bowels before starting pranayama. The best time for practice is the early morning, preferably before sunrise when the pollution is at its lowest level, and the body and brain are relatively relaxed. However, if morning is unsuitable, pranayama may be practiced after sunset, when the air is cool and pleasant. The place suitable for all kinds of Yoga must be clean and

calm. The practice of pranayama should be preferably carried out 3 hours after taking solid food and 1 hour after taking liquid food.

We chose our subjects in the age group 16-19 years, who practiced Pranayama, in Padmasana posture at evening time (6 p.m. to 7 p.m.) after attending their course classes. This age group was specifically selected for the reason that the newly found energy in this age is often not utilized judiciously. Practicing the pranayama helps in channelising this enormous energy [10].

The subjects practiced Pranayama for three months regularly.

There are various techniques of Pranayama but we applied the technique of Nadi Shodhan Pranayama on the subjects. This Pranayama is one of the simplest exercises which require no pre-requisite and the technique followed by the subjects is given in the subsequent section.

In NSP, the subjects are made to sit down in a comfortable place assuming a cross legged position on a mat. They are instructed to sit erect, remain calm and close their eyes. The breathing process starts by closing the right nostril with the right thumb, followed by inhaling slowly through the left nostril. After complete inhalation, the left nostril is to be pressed with the ring finger of the right hand and close the left nostril. Then the right nostril is opened to exhale slowly. After complete exhalation, the breath is inhaled through the right nostril. It is followed by closing the right nostril by pressing it with the right thumb. Finally the left nostril is opened to breathe out slowly. This explicitly described process is called one round of Nadi Shodhan Pranayama or Anuloma Viloma Pranayama (Alternate Nostril Breathing) and is depicted in figure P2. This is to be continued for 10 -15 rounds [11], [12], [13] and [14].

Analysis

We chose a group of twenty subjects who practised Pranayama techniques for three months. The different characteristics such as Feeling Healthy, Memory Recall, Mental Stress Relief and Physical Relaxation were observed on the basis of the self introspection i.e. subjective observation by the person practicing these techniques. These were recorded using a well established Likert's five-point psychometric scale [15]. In this scale the subject is asked to self introspect his status of his trait in term of percentage. This percentage is taken as 10%, 30%, 50%, 70% and 90%. For example, if the person is feeling too weak physically and thinks he is terribly sick, he may choose 10% as his status of health. If the subject feels that he is somewhat

sick he may choose 30% on this Likert's scale. Conversely, if he feels that he is in perfect state of health, he may tick at 90%, while for more than average health, he may choose 70%. For average health condition, he may choose 50%. Similarly the choice is made for all other traits. It is a standard practice with Likert's five-point psychometric scale to simplify the analysis by clubbing the 10% and 30% categories and considering it as 20% and assumed as Low scale, 50% is retained unchanged and assumed at Medium Scale while and the values of 70% and 90% are clubbed and considered as 80%, assumed as High Scale.

The results are shown in Table 1.

The variation of these characteristics (before and after practicing Nadi Shodhan Pranayama techniques) has been shown in the figures (1 - 4). These effects are mentioned as follows.

Results & Discussion

- A NSP technique was applied on a set of 20 engineering students. None of the students reported any decline in the status of health after this Pranayama. 15 out of these 20 students reported better in Feeling Healthy after this exercise. It can therefore be concluded 75% students (15 out of 20) improved in the Feeling Healthy level after Pranayama. Figure {1(a) - 1(b)}
- Similarly 80% of the students (16 out of 20) experienced better state of Memory Recall after Pranayama. None of the students reported any decline. Figure {2(a) - 2(b)}

- Mental Stress level of 75% students (15 out of 20) decreased after Pranayama. None of the students reported any decline. Figure {3(a) - 3(b)}
- State of Physical Relaxation of 90% students (18 out of 20) improved after Pranayama. None of the students reported any decline. Figure {4(a) - 4(b)}

After Pranayama, the low scale values of the Memory Recall and Physical Relaxation characteristics have been reduced to nil, whereas high scale values of Feeling Healthy, Memory Recall and Physical Relaxation have been drastically increased (Table 2, 3 and 5). The results for the Mental Stress characteristics are highly encouraging. Table 4 indicates drastic shift of Mental Stress from high scale to low scale after practice of Nadi Shodhan Pranayama.

The results obtained above are for the prescribed optimal technique of alternate nostril breathing. A common query that was posed by many participants to the authors was, what if it is not practically possible to attain these conditions?

It is hereby assured that the practitioners shall get some stress relief, even if he is not sitting in cross-legged position, or has skipped the gap time i.e. has eaten some solid or liquid food to quench his hunger and thirst. The result will be positive though not to the extent documented above, even if this technique is applied for shorter duration of time while traveling or sitting in office chair.

Table 1. Characteristics analysis before and after practicing Nadi Shodhan Pranayama (NSP)

S No.	Student Codes	Likert's scale of Feeling Healthy		Likert's scale of Memory Recall		Likert's scale of Mental Stress		Likert's scale of Physical Relaxation	
		Before	After	Before	After	Before	After	Before	After
1.	S1	50	70	50	70	50	10	70	90
2.	S2	50	70	50	70	70	50	30	70
3.	S3	30	90	50	50	70	30	50	70
4.	S4	50	90	50	70	70	70	50	70
5.	S5	70	90	50	70	70	70	50	70
6.	S6	50	90	50	70	70	70	30	90
7.	S7	70	90	70	90	50	10	50	90
8.	S8	70	70	30	50	50	50	50	70
9.	S9	30	70	30	70	70	30	70	90
10.	S10	50	90	50	50	70	50	50	50
11.	S11	70	90	50	70	70	70	50	50
12.	S12	50	50	50	50	70	30	30	50

S No.	Student Codes	Likert's scale of Feeling Healthy		Likert's scale of Memory Recall		Likert's scale of Mental Stress		Likert's scale of Physical Relaxation	
		Before	After	Before	After	Before	After	Before	After
13.	S13	30	50	50	70	50	30	30	50
14.	S14	30	30	50	70	70	70	70	90
15.	S15	50	50	30	50	30	10	30	50
16.	S16	30	50	50	70	50	30	30	50
17.	S17	30	70	50	70	50	30	30	50
18.	S18	30	70	30	70	90	50	30	50
19.	S19	50	50	50	50	70	30	30	70
20.	S20	50	70	30	70	50	30	30	90

Feeling Healthy; Memory Recall; Mental Stress Relief; Physical Relaxation.

Table 2

Scale of Feeling Healthy	Before Pranayama	After Pranayama
Low scale	35% (7 out of 20)	05% (1 out of 20)
Medium scale	45% (9 out of 20)	25% (5 out of 20)
High scale	20% (4 out of 20)	70% (14 out of 20)

Table 3

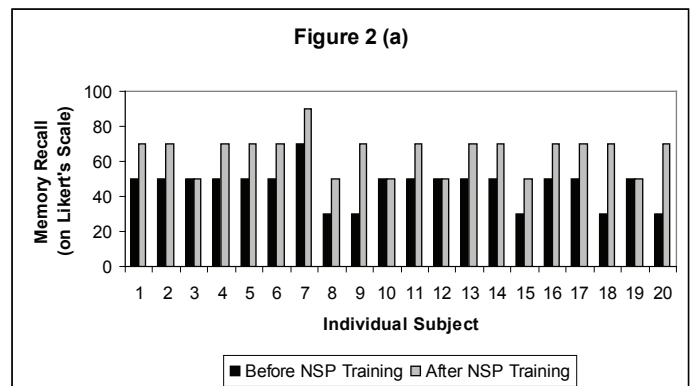
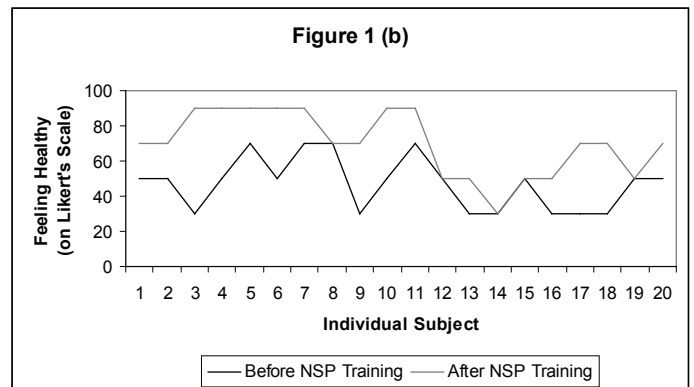
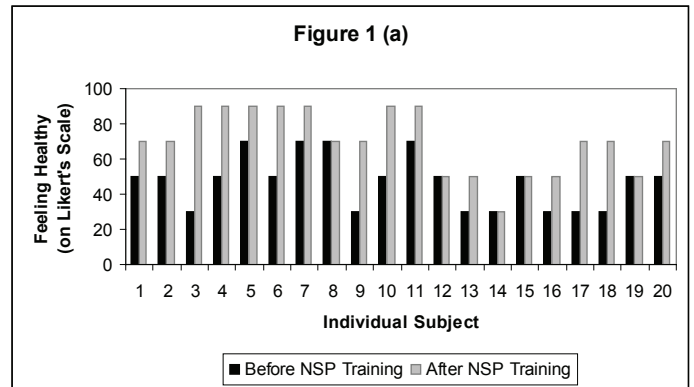
Scale of Memory Recall	Before Pranayama	After Pranayama
Low scale	25% (5 out of 20)	Nil
Medium scale	70% (14 out of 20)	30% (6 out of 20)
High scale	05% (1 out of 20)	70% (14 out of 20)

Table 4

Scale of Mental Stress	Before Pranayama	After Pranayama
Low scale	05% (1 out of 20)	55% (11 out of 20)
Medium scale	35% (7 out of 20)	20% (4 out of 20)
High scale	60% (12 out of 20)	25% (5 out of 20)

Table 5

Scale of Physical Relaxation	Before Pranayama	After Pranayama
Low scale	50% (10 out of 20)	Nil
Medium scale	35% (7 out of 20)	40% (8 out of 20)
High scale	15% (3 out of 20)	60% (12 out of 20)



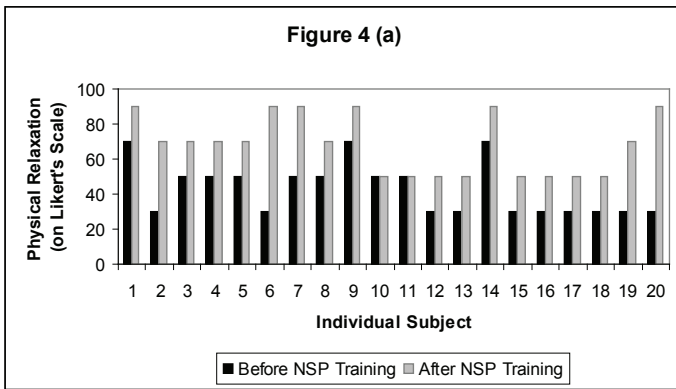
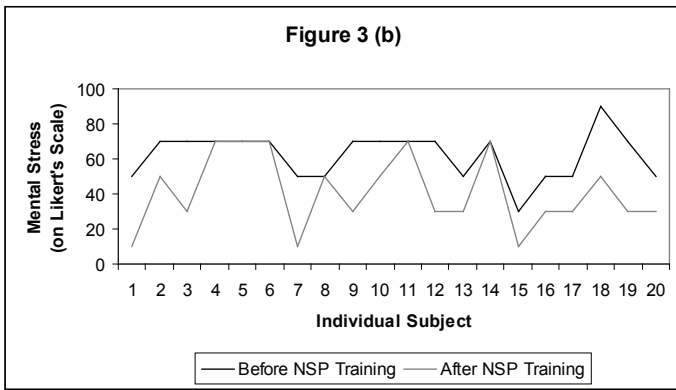
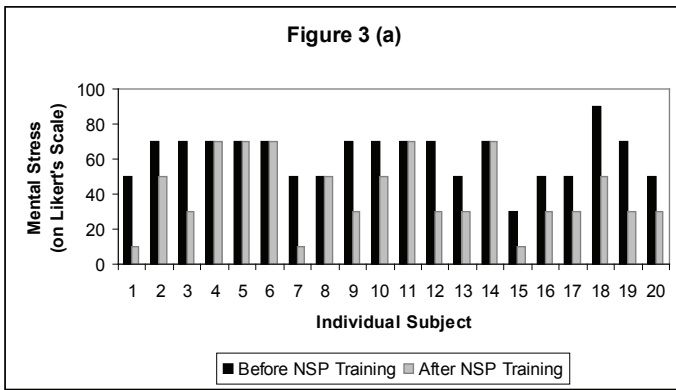
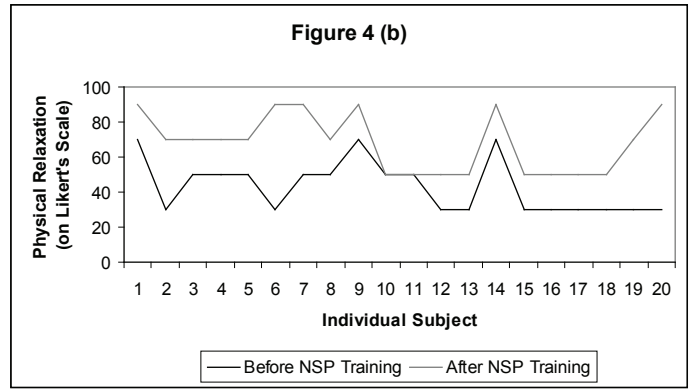
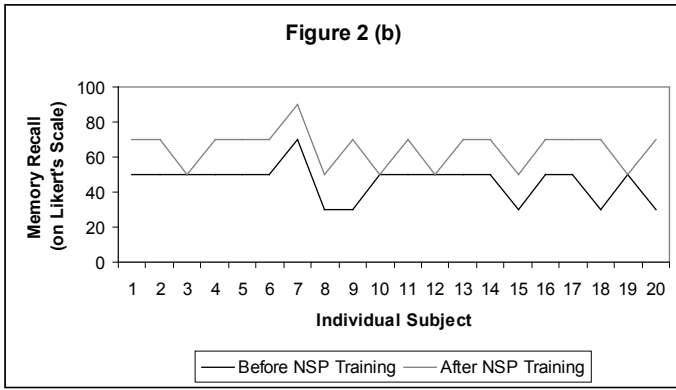


Figure P1

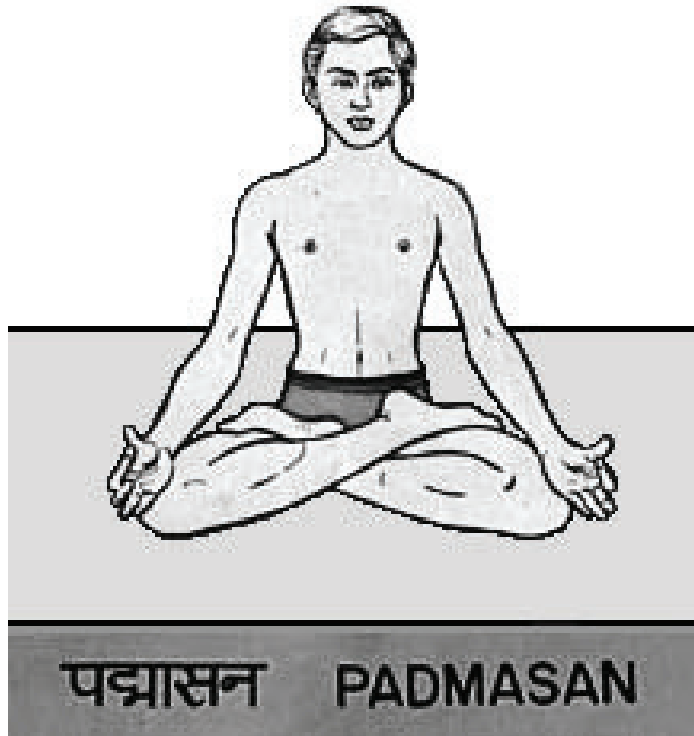


Figure P2



NADI SHODHAN PRANAYAMA POSTURE

The observed results are analyzed using t-test for knowing the statistical significance and presented in tables 6-9. The t-test is used to identify the significant difference in the means of two samples, namely before and after practising nadi shodhan pranayama.

Table 6

t-Test		
Feeling Healthy	Before Meditation	After Meditation
Mean	47	70
Variance	222.1052632	336.8421053
Observations	20	20
Pearson Correlation	0.461811804	
Hypothesized Mean Difference	0	
df	19	
t Stat	-5.877033172	
P(T<=t) two-tail	1.16684E-05	
t Critical two-tail	2.09302405	

Table 7

t-Test		
Memory Recall	Before Meditation	After Meditation
Mean	46	65
Variance	109.4736842	121.0526316
Observations	20	20
Pearson Correlation	0.365758478	
Hypothesized Mean Difference	0	
df	19	
t Stat	-7.024622676	
P(T<=t) two-tail	1.09232E-06	
t Critical two-tail	2.09302405	

Table 8

t-Test		
Mental Stress	Before Meditation	After Meditation
Mean	62	41
Variance	185.2631579	441.0526316
Observations	20	20
Pearson Correlation	0.618651309	
Hypothesized Mean Difference	0	
df	19	
t Stat	5.687824835	
P(T<=t) two-tail	1.75242E-05	
t Critical two-tail	2.09302405	

Table 9

t-Test		
Mental Stress	Before Meditation	After Meditation
Mean	62	41
Variance	185.2631579	441.0526316
Observations	20	20
Pearson Correlation	0.618651309	
Hypothesized Mean Difference	0	
df	19	
t Stat	5.687824835	
P(T<=t) two-tail	1.75242E-05	
t Critical two-tail	2.09302405	

t-test (dependent Sample)

t-test is used to identify the significant difference in the means of two dependent samples.

Null Hypothesis (Ho :)

There is no significant difference in the means of the parameters of Feeling Healthy, Memory.

Recall, Mental Stress and Physical Relaxation before and after practising nadi shodhan pranayama.

Alternative Hypothesis (Ha :)

There is a significant difference in the means of the parameters of Feeling Healthy, Memory.

Recall, Mental Stress and Physical Relaxation before and after practising nadi shodhan pranayama.

Results of t-test

From the Table 6-9 the t-critical value for all the parameters is 2.093 and the standard value at 1% significance level (99% confidence level) is 1.96. For interpretation, if the t-calculated value is less than the standard value then Ho : is accepted. But in this case t-calculated value is more than the standard value at both the levels. So here Ho : is rejected and Ha : is accepted i.e. there is a significant difference in the means of the parameters of Feeling Healthy, Memory Recall, Mental Stress and Physical Relaxation among the students before and after practising nadi shodhan pranayama i.e. the technique of practising nadi shodhan pranayama is proved to be useful for the subjects. Also it is observed from the results of t test, mean value of Feeling Healthy is shifted from 47 to 70, Memory Recall is shifted from 46 to 65, Mental Stress is shifted from 62 to 41 and Physical Relaxation is shifted from 43 to 68 before and after the test. The average value in this case is 50 since is the lowest and 90 is the highest value, while recording the responses. It means the content of Mental Stress declines where as the level of Feeling fHealthy, Memory Recall and Physical Relaxation rises after practising nadi shodhan pranayama.

Conclusion

Yoga is a proven technique for bringing desirable changes in behavioral traits leading to wellbeing. The present study affirmatively proves the effect of practice of Nadi Shodhan Pranayama (NSP) in enhancing Feeling Healthy, Memory Recall, Mental Stress Relief and Physical Relaxation. It was observed that 75% of the subjects gained in terms of Feeling Healthy, 80% in terms of Memory Recall, 75% in terms of Mental Stress Relief and 90% in terms of Physical Relaxation. NSP improves the traits of Feeling Healthy,

Memory Recall and Physical Relaxation characteristics. It reduces Mental Stress level, which is highly contributing factor to disturb wellbeing of anyone. Also the higher stress level leads to many diseases. The regulation and control of above traits is highly significant to enhance the wellbeing of Engineering students. The authors want to spread the message through this research paper that to maintain good health, to get rid of diseases and lead a fulfilling life, which is the basic necessity at all levels of our society, NSP is a very effective technique. It may be safely concluded that NSP has immense potential of increasing the wellbeing of its practitioners irrespective of their age. It may become a universal tool for making this globe a happier place.

References

- [1]“TCM and Acupuncture Health Info” available at <http://tcm.health-info.org/>
- [2]Dr. Padmasiri de Silva “The Psychology of Emotions in Buddhist Perspective” Sir D. B. Jayatilleke Commemoration Lecture Colombo, 1976 Buddhist Publication Society Kandy • Sri Lanka The Wheel Publication/237
- [3]Mandlik Gandhar “Pranayama - A Science of Breathing (I) “available at www.yogapoint.com/pranayama/pranayama_science_1.htm
- [4]Upadhyay Avnish K, Balkrishna Acharya, Upadhyay Ruchita T. Effect of Pranayama [Voluntary Regulated Yoga Breathing] and Yogasana [Yoga Postures] in Diabetes Mellitus (DM) A Scientific Review. Journal of Complementary and Integrative Medicine, 2008; 5: 1:1:3. Available at:<http://www.bepress.com/jcim/vol5/iss1/3>
- [5]Raghuraj P, Ramakrishnan A G, Nagendra H R, Telles Shirley. Effect of Two Selected Yogic Breathing Techniques on Heart Rate Variability available at http://www.healthandyoga.com/html/research_papers/ets/om.asp
- [6]Jerath R, Edry J, Barnes V, Jerath V. Physiology of long pranayamic breathing: Neural respiratory elements may provide a mechanism that explains how slow deep breathing shifts the autonomic nervous system Medical Hypotheses, 2006; 67: 1:3:566-571.
- [7]Joshi Anurag, Singh Mandeep, Jindal Ritu, Parkash Jai. Role of Acoustic Meditation in Stress Management An Analysis. International Journal of Management Sciences, 2008; 4:2 (Dec. 2008) 56-67.

- [8]Joshi Anurag, Joshi Sunil, Singh Mandeep, Kaur Sukhwinder. STRESS A BANE – YOGA THE CURE in National Seminar on Ayurveda, 2009; 20-24.
- [9]Dr. Arun Kumar SR “Effect of Nadi Shodhana Pranayama on Autonomic functions among healthy young school children in the age group of 11-16 years” Dissertation of Master Degree in Physiology, Submitted To The Rajiv Gandhi University Of Health Sciences, Bangalore (Karnataka), India (2006).
- [10]“Secret of Eternal Youth” available at <http://www.ashram.org/doc/SecretOfEternalYouth.pdf>
- [11]Nadi Shodhan Pranayama <http://www.yogapoint.com/info/pranayama.htm>
- [12]“Nadi Shudhi” available at <http://www.healthandyoga.com/html/pran/nadishudhi.htm>
- [13]“Breathing Exercise (Pranayama) - Alternate Nostril (Anuloma Viloma)” {Demo Pictures of One Round of Anuloma Viloma (Alternate Nostril Breathing)} available at <http://www.abc-of-yoga.com/pranayama/basic/viloma.asp>
- [14]Video demonstration of “Alternate Nostril Breathing (Nadi Sodhana)” available at http://www.metacafe.com/watch/735686/alternate_nostril_breathing_nadi_sodhana/
- [15]Walia J.S. “Chief Methods of Psychology”, Manual of Psychology, 1st ed. Paul Publishers, Jalandhar (Punjab) India, 2002.16-17.

An Effective Neuroprotective Treatment in Ischemic Stroke and Cerebral Trauma with Low Doses of L-Arginine, Lamotrigine and Tianeptine

Héctor Jara (MD)

Department of Neurochemistry, Neuropharmacology and Neuroimmunology, Instituto de Medicina Experimental. Faculty of medicine, Universidad Central de Venezuela, Caracas, Venezuela; Neuropharmacology Institute at the “Centro Clínico Profesional Caracas”, Caracas, Venezuela.

[Email: hectorjara2000@hotmail.com](mailto:hectorjara2000@hotmail.com)

Francisco García (MD)

Neuropharmacology Institute at the “Centro Clínico Profesional Caracas”, Caracas, Venezuela.

Franky Torres (MD)

Department of critical care medicine, Hospital “Domingo Luciani” IVSS, Caracas, Venezuela.

Fuad Lechin (MD,PhD)

Department of Neurochemistry, Neuropharmacology and Neuroimmunology, Instituto de Medicina Experimental. Faculty of medicine, Universidad Central de Venezuela, Caracas, Venezuela.

Abstract

In stroke and cerebral Trauma the damaged neurons release Aspartate and Glutamate that contribute to neuronal death (excitotoxic cell death). However, physiological levels of NMDA receptor activity can promote neuronal survival and resistance to trauma, which explain why a large number of neuroprotective agents development for stroke have failed to show positive effects in Phase III trials in ischemic stroke. The levels of free serotonin (f-5HT) increase in thrombotic events, worsening the platelet aggregation and the arteriolar vasoconstriction. Endothelium-derived Nitric Oxide (eNO) is reduced in ischemic stroke. The objective of this study was to evaluate the neuroprotective effects of Lamotrigine (a glutamate release inhibitor), Tianeptine (reducer of f-5HT levels) and L-Arginine (precursor of eNO) at low doses. We performed a controlled study with 49 patients with cerebral trauma and 25 patients with ischemic stroke. We compared the sample groups with control groups that received conventional treatment. To evaluate the disease progression we used the Glasgow Scale and NIHSS. The results were analyzed by F statistical test and Student's t test with $p \leq 0.05$. The results show that patients with stroke and brain trauma are benefited with this new neuroprotective treatment (p value < 0.05) with lower rates of neurologic complications. **Key Words:** Glutamate, NMDA-receptor signalling, free serotonin, endothelial Nitric Oxide, Agmatine, Pro-survival signal, Neurotoxicity.

Cerebrovascular Diseases (CD), the third leading cause of death in developed countries after heart diseases and cancer, has an overall prevalence of 795 per 100,000 and are a major cause of disability. Two thirds of stroke survivors suffer from residual neurological deficits and have to cope with chronic motor and language dysfunctions. So far, we have very limited effective therapies in spite of intensive research efforts and numerous clinical trials (1,3,21). A stroke is the acute neurologic injury occurring as a result of several pathological processes involving the blood vessels of the brain. Normal brain function requires continuous

supply of oxygenated blood. Reduction in blood flow may interfere with brain functions, but the brain can remain viable for more prolonged periods (37), for example after a stroke patients often recover partially or completely, suggesting that focal areas of the brain can remain functionless and ischemic for hours, even days. This has led to the notion of an ischemic zone (penumbra or halo) that surrounds the infarct area and could progress up to neuronal death or is potentially salvageable if ischemia can be reversed (11,37). There are also secondary phenomena that may contribute to neuronal death such as excitatory amino acids

(Glutamate and Aspartate) released by damaged neurons (Excitotoxicity), cerebral oedema and alterations in local blood flow owing to endothelial responses. (5,11,37) A small number of ischemic stroke patients are eligible for thrombolytic therapy with tissue plasminogen activator (t-PA) but this has to be administered within 3 hours of the ischemic event (5). However, several potential side effects have been reported.

There is much information concerning the early use of neuroprotection in ischemic stroke and brain trauma. There is a lot of evidence indicating that NMDA-receptors (N-Methyl D-Aspartate-receptors) and voltage-dependent calcium channels could be one of the triggers of neuronal injury after ischemic stroke (Xion et al, 2004) (51). Choi (1988) (7) and others (5,27) first worked with drugs that block calcium influx into ischemic cells; either conventional calcium channel blockers or Glutamate receptor blockers have had variable success in patients with stroke and in animal models. They demonstrate that the early use of NMDA-Blockers (N-Methyl D-Aspartate-Blockers) could prevent the progression of the stroke in patients who undergo cardiac catheterization. Although these results were promising, there are a lot of researchers who used these drugs with poor results (8,27,28). Several drugs that seemed promising in experimental studies or in small trials (including Glycine-NMDA-Receptor-Antagonists) have proved ineffective in phase 3 trials (8,21,28,29). The observed lack of efficacy of these drugs may be due to delays in the initiation of the treatment; however the dichotomy of NMDA-receptor signalling is a more plausible explanation (40). In pathological scenarios such as ischemia, calcium influx through the NMDA-receptor is a key mediator of cell death. Nevertheless, physiological levels of NMDA-receptor activity can promote neuronal survival and resistance to trauma, and play important roles in synaptic plasticity. (40). There is evidence that physiological synaptic NMDA-receptor activity exerts a neuroprotective effect. It may play a role in promoting recovery and preventing delayed neuronal loss in the penumbra (17).

The antiepileptic drug Lamotrigine is a phenyltriazine derivative that acts by stabilizing voltage-sensitive sodium channels in a usage-dependent manner, preventing glutamate and aspartate release and reversibly blocking excitatory neurotransmission. In previous studies has been demonstrated to be effective for hypoxic-ischemic brain damage in focal and global stroke models (9,12,45,50,53) as well as combination therapy for the patient with ischemic stroke (6). Low doses of Lamotrigine could avoid

CNS-adverse events and others side effects. On the other hand these doses could prevent global NMDA antagonists that may block NMDA-receptor-activated pro-survival signal triggered in response to an ischemic challenge. This dichotomy of NMDA-receptor signalling (16) means that any anti-excitotoxic strategy that interferes with NMDA-receptor signalling should be assessed to determine its effects on NMDA-receptor pro-survival signalling.

Total circulating serotonin includes platelet-serotonin (p-5HT) + free-serotonin (f-5HT) in the plasma. There is a lot of evidence that f-5-HT plasma levels augment in cerebrovascular diseases and trigger platelet aggregation enhancing the endothelium-dependent vasoconstriction, (23,31) which worsens the ischemia. The fact that a small dose of oral Tianeptine, a drug that enhances serotonin uptake, reduces f-5HT in the plasma (23,24,25) motivates us to include it in the neuroprotective treatment. On the other hand we have been using Tianeptine during the last 14 years to reduce the f-5HT plasma levels in asthmatic patients with successful results. (25)

In literature it has been proved that Nitric Oxide (NO) induces vascular smooth muscle relaxation (Palmer et al, 1987) (38), and is synthesized in the endothelial cells from L-Arginine (Palmer et al, 1988) (39) by the enzyme nitric oxide synthase (NOS), and it modulates a wide variety of neural, cardiovascular, endocrinologic and humoral processes. The endothelial NO release is reduced in stroke due to endothelial factors particularly in the cerebral vasculature. (31,46). On the other hand Agmatine, formed by the decarboxylation of L-arginine by arginine decarboxylase, has been shown to be neuroprotective in trauma and ischemia models (19,36).

The current treatment for stroke relies on the use of thrombolytic agents, which are of demonstrable value only if delivered within three hours after the onset of the stroke. Although potential side effects must be considered, a neuroprotective treatment that reduces the Glutamate releases avoiding to block the pro-survival effects of NMDA-receptor activity, enhances the endothelial NO and reduces f-5HT could be an attractive option for new stroke and cerebral trauma therapies. In the present study we use low doses of Lamotrigine (Inhibitor of glutamate and aspartate release) + L-Arginine (An endothelial NO precursor) + Tianeptine (drug that reduces the f-5HT) in stroke and cerebral trauma patients to reverse the progression of cerebral ischemia toward cell death by necrosis or apoptosis.

Material and Methods

Subjects: 131 patients (78 male-53 female) between 18-85 years old conformed the eligible patients group. The institutional review board of each participating centre approved the protocol.

Inclusion Criteria

- Clinical diagnosis of stroke or cerebral trauma.
 - Onset of symptoms to time administration:
- Patients who had suffered the stroke less than 48 hours before treatment.

In cerebral trauma patients who did not respond to conventional therapy (Manitol, Somazina and conventional treatment) during at least 4 days.

- CT Scan showing cerebral ischemia or oedema/haemorrhage (cerebral trauma).
- Age > 30-year-old patients with stroke
- Age >18 year-old patients with cerebral trauma.
- In Stroke patients: National Institute of Health Stroke Scale (NIHSS) score 4-20

Exclusion Criteria

- Glucose > 200 mg/d
- Fever
- Hypercalcemia
- Minor stroke symptoms and TIA or NIHSS less than 4 points.
- Extensive stroke with Glasgow of 3 points and NIHSS upper 20 points.
- Haemorrhage Stroke.

Baseline clinical assessment

At baseline, details of the medical history were established by interview and consultation of medical notes. (See table 2) Patients with stroke were examined neurologically and classified as total anterior circulation syndrome (TAC), partial anterior circulation syndrome (PAC), lacunar syndrome (LAC) and posterior circulation syndrome (POC) using the Oxfordshire Community Stroke Project (OCSP) Classification (4) (See table 3). Cerebral MRI confirmed all cases within 3-7 days since the stroke occurred. Neurological deficit was scored using the National Institute for Health Stroke Scale (NIHSS). The NIHSS was included because it is the most widely used stroke scale.

The Glasgow Scale was used to evaluate the coma state in all the groups.

The Group A (Sample group)

74 patients (50 Male-24 Female) sub divided in two sub-groups:

A.1) 49 patients with cerebral trauma (39 male-10 female-18 to 60 years old)

A.2) 25 patients with stroke (11 male-14 female-60 to 85 years old)

All the A.1 group patients had cerebral trauma (Hospital Rafael Medina Jimenez-La Guaira-Venezuela and Hospital Domingo Luciani IVSS-Caracas-Venezuela) with traumatic haemorrhage and or cerebral oedema. Most of them with skull fractures and coma categorized as "Diffuse axonal injury". (See table 1) The study began with those patients who did not respond to conventional therapy during at least 4 days from the entrance to the intensive care unit.

To evaluate the disease progression we used the Glasgow Scale at the beginning of the study and at the discharge of the intensive care unit. In order to maintain the same clinical conditions we compared the A.1 patients group to a 32-patients control group (20 male-12 female) with cerebral trauma who did not respond to conventional treatment after 4 days in the intensive care unit.

All the A.2 group patients (25 patients) had cerebral ischemia (Stroke) (Centro Clinico Profesional Caracas-Caracas-Venezuela and Policlínica Las Mercedes- Caracas-Venezuela). 14 male-11 female. The study began in the first 48 hours since the onset of the symptoms. To evaluate the disease progression we used the Glasgow Scale and the NHI Stroke Scale at the beginning of the study and then every 24 hours. In order to maintain the same clinical conditions we compared the A.2 group patients with a control group of 25 patients sex-age-matched (8 male-17 female), they received conventional treatment. The study began in July-2008 and was completed in December-2010.

If the f-5HT and Glutamate worsen the ischemic and neuronal injury (cerebral trauma) and enhance the progress up to neuronal death via apoptosis or necrosis, then with a treatment that enhances the eNO, reduces the f-5HT and reduces partially the activation of the NMDA-receptors, the reduction of the progression of cerebral ischemia and the changes of the clinical patient's evolution could improve.

According to the above-mentioned statement the eligible group patients (A.1 and A.2) were treated orally with 12.5 mg of Tianeptine (drug that reduces f-5HT), 25

mg of Lamotrigine (Inhibitor of Glutamate) and 500 mg of L-Arginine (An NO precursor) twice daily during 10 days and then once daily.

Protocol for Collecting Data

4 different physicians carried out the experimental procedures. Two of them with the A.1 group (specialists in critical care medicine). The remaining two with the A.2 group (specialists in internal medicine).

- The Glasgow scale and the NIHSS were used in the sample group and control group with stroke.
- The Glasgow scale was used in the sample group and control group with cerebral trauma.

Statistical procedures used

- The data were summarized by means and percent.
- We used numerical and graphical techniques throughout the study.
- Comparisons of values between control group and treated group were analyzed by F statistical test and Student's t test with $p \leq 0.05$.

Table 1. Baseline characteristics of patients with cerebral trauma based on treated Group (A1) and Control

Characteristics	A1 Group Cerebral Trauma (N= 49)	Control Group (Cerebral Trauma) (N= 32)
Age-yr	39,14	43,5
Average Age by Sex in Male (yr.)	39,38	38,9
Average Age by Sex in Female (yr.)	38,2	51,16
Male sex (%)	79,59	62,5
Female sex (%)	20,41	37,5
DSF No. (%)	8 (16,32)	5 (15,62)
BSF No. (%)	4 (8,16)	3 (9,37)
DAI No. (%)	9 (18,36)	3 (9,37)
CE No. (%)	31 (63,26)	13 (40,62)
H/H No. (%)	28 (57,14)	17 (53,12)
CSF No. (%)	3 (6,12)	1 (3,12)
DSF = Depressed Skull Fracture	BFS = Basilar Skull Fracture	DAI = Diffuse Axonal Injury
CE = Cerebral Edema	H/H =Hemorrhage/ Hematoma	CSF = Comminuted Skull Fracture

Note: All patients showed two or more findings

Table 2. Baseline characteristics of ischemic stroke patients based on treated Group (A2) and control Group.

Characteristic	A2 Group N = 25	Control Group N = 25
Age-yr	74.6	75.2
Male sex Age-yr	76.9	75.3
Female sex Age-yr	72.2	75.17
Male sex No. (%)	11 (44%)	8 (32%)
Female sex No. (%)	14 (56%)	17 (68%)
Diabetes Mellitus No. (%)	11 (44%)	13 (52%)
Arterial Hypertension No. (%)	10 (40%)	13 (52%)
Current Smoker No. (%)	2 (8%)	2 (8%)
Prior cardiovascular disease No. (%)	3 (12%)	2 (8%)
Hyperlipidemia	11(44%)	9 (36%)
Prior stroke or TIA No. (%)	2 (8%)	1 (4%)

Table 3. Details of the stroke after the OCSF classification and cerebral MRI

Details of the Stroke (After MRI)		
Details of the Stroke	A2 Group (Ischemic Stroke) (No. = 25)	Control Group (Ischemic Stroke) (No. = 25)
TACI	5 (20%)	4 (16%)
PACI	6 (24%)	8 (32%)
LACI	9 (36%)	8 (32%)
POCI	4 (16%)	5 (20%)
Unclassified (after MRI)	1 (4%)	0
HC	0	2 (8%)

TACI = Total Anterior Circulation Infarction

PACI = Partial Anterior Circulation Infarction

LACI = Lacunar Infarction

POCI = Posterior Circulation Infarction

HC = Hemorrhagic Conversion in Ischemically Infarcted Areas

Unclassified = The MRI Showed old frontal Infarction at the 7 days, but the OCSF Classification and the Scan at 24 hours showed TACI

Results

A.1 Group (Cerebral Trauma)

Age Average: 39.14 years old

Sex:

- Female: 20.41 %
- Male: 79.59 %

Glasgow Scale Average (Pre-Treatment): 5.33 points

Glasgow Scale Average (Post-treatment): 11.29 points

(See Fig. 1)

Improved Patients: 40/49 patients (81.63 %)

Patients with Glasgow Scale improvement but without coming out of the coma state: 9 patients (18.37 %)

Mortality: 8.16 % and Sepsis: 8.16 %. (See Fig. 2)

81.63 % of the A.1 group improved the Glasgow Scale in 5.8 points or more in two weeks. (See fig.1)

18.36 % of the A1 group improved the Glasgow Scale at least in 3 points but they remained in the coma state.

The time average of Glasgow recuperation in at least 5.8 points was two weeks in 81.63 % of the patients and the time average of partial recuperation was three weeks in 18.36 % of the patients.

Control Group (Cerebral Trauma)

Age Average: 43.5 years old

Sex:

- Female: 37.5 %
- Male: 62.5 %

Glasgow scale average (pre-treatment): 5.77 points

Glasgow scale average (post-treatment): 7.75 points

The Glasgow scale average pre treatment was 5.77 points and post-treatment (until 4 weeks after) 7.75 points.

Improved Patients: 5/32 patients (15.62 %)

Patients with Glasgow Scale improvement but without coming out of the coma state: 6/32 patients (18.75 %)

Patients without Glasgow Scale improvement: 21/32 patients (65.6 %)

Mortality: 21.87 % and Sepsis: 18.75 % (See Fig 2)

A.2 Group (Cerebral Stroke)

Age Average: 74.60 years old

Sex:

- Female: 14 patients (56 %)
- Male: 11 patients (44 %)

Glasgow Scale Average (Pre-Treatment): 8.8 Points

Glasgow Scale Average (Post-Treatment): 15 Points

NIH Stroke Scale Average (Pre-treatment): 16.4 Points

NIH Stroke Scale Average (Post-treatment): 3 points

(See Fig. 4 and 5)

Time of Complete Recovery based in Glasgow scale in the A.2 group

1. Patients who started the treatment within the first 3 hours since the onset of the symptoms: 8 patients (32 %). The mean time of complete recovery was 58 hours.
2. Patients who started the treatment within the first 6-36 hours (mean time: 12.44 hours) since the onset of the symptoms: 17 patients: 68 %. The mean time of complete recovery was 5.7 days.

Time of Complete Recovery based in NIHSS in the A.2 group

1. Patients who started the treatment within the first 6 hours since the onset of the symptoms: 15 patients (60 %). The mean time of the complete recovery was 51.2 hours.
2. Patients who started the treatment within the first 8-36 hours since the onset of the symptoms: 10 patients (40 %). The mean time of complete recovery was 5.28 days.

Control Group (Cerebral Stroke)

Age Average: 75.24 years old

Sex:

- Female: 17 (68 %)
- Male: 8 (32 %)

Glasgow Scale Average Pre-Conventional Treatment: 9.48 points

Glasgow Scale Average at 72 hours: 12.4 points. (See Fig. 8)

NIH Stroke Scale Average pre-treatment: 16.84 points

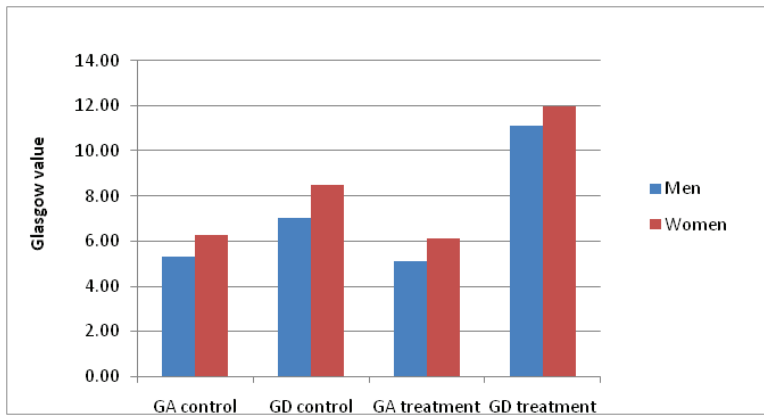
NIH Stroke Scale Average post-treatment (24 hours): 16.24

NIH Stroke Scale Average post-treatment (48 hours): 15.72 points

NIH Stroke Scale Average post-treatment (one week): 12.72 points

NIH Stroke Scale Average post-treatment (two weeks): 10.12 points (See fig. 6)

Fig. 1. Comparison by sex of the Glasgow Scale at admission (GA) and at discharge (GD) in the AI Group (cerebral trauma) and the Control Group.



In all cases P value < 0.05

Fig. 2 Complications in cerebral trauma (Sample group and control)

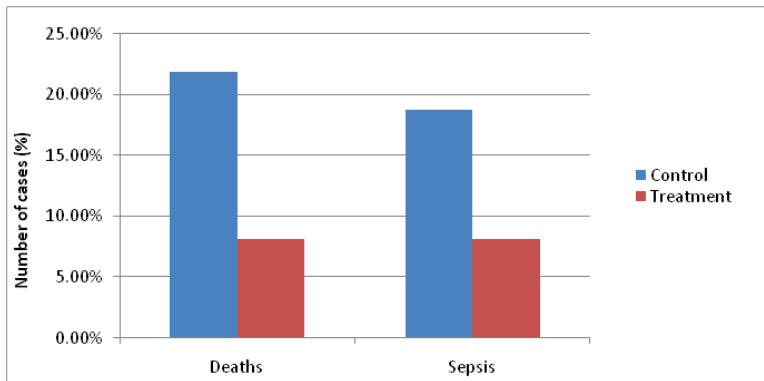
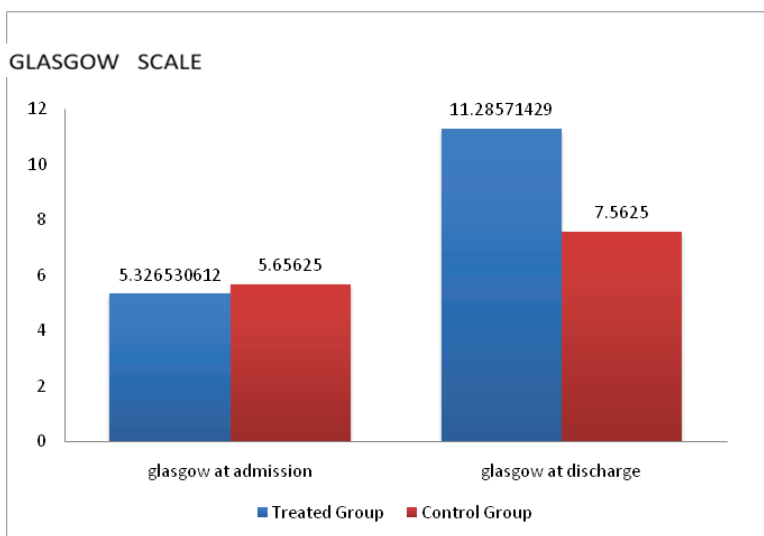
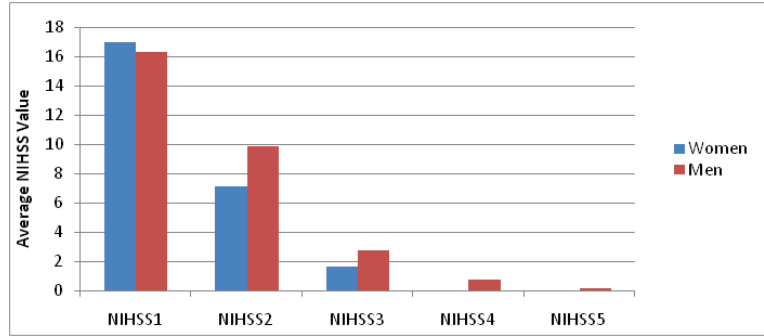


Fig 3. Glasgow comparison between Treated and Control Group at admission and discharge in cerebral trauma



P value < 0.05

Fig. 4. Evolution of the NIHSS in women and men (A2 Group: Stroke Group)

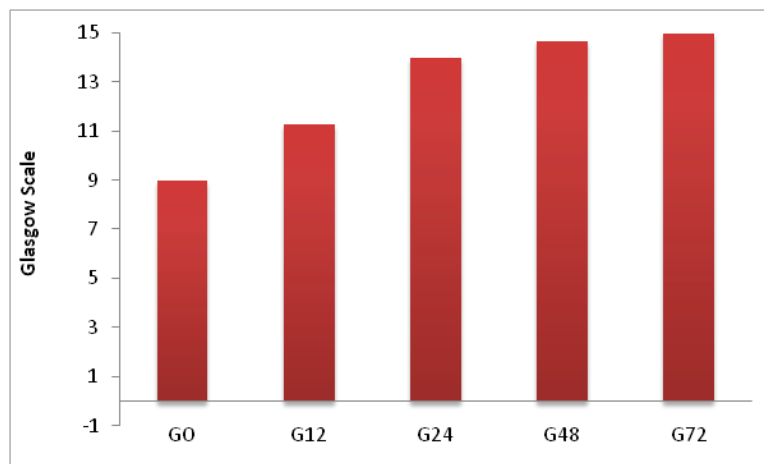


NIHSS: National Institute Health Scale Score

NIHSS1: At admission; NIHSS2: At 24 hours; NIHSS3: At 48 hours; NIHSS4: At one week; NIHSS5: At two weeks.

P value < 0.05 in NIHSS2, NIHSS3, NIHSS4 and NIHSS5.

Fig. 5. Glasgow evolution in the A2 Group (Ischemic Stroke)



G0: Glasgow at admission; G12: Glasgow after 12 hours; G24: Glasgow after 24 hours; G48: Glasgow after 48 hours; G72: Glasgow after 72 hours.

P value < 0.05 in G12, G24, G48 and G72.

Fig. 6. Evolution of NIHSS in Control Group by sex (Ischemic Stroke)

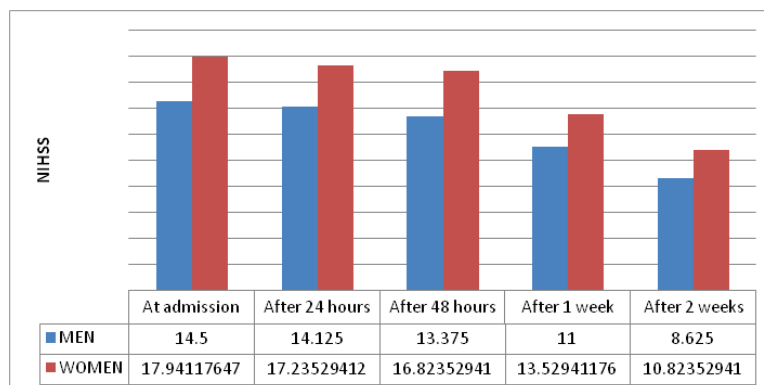
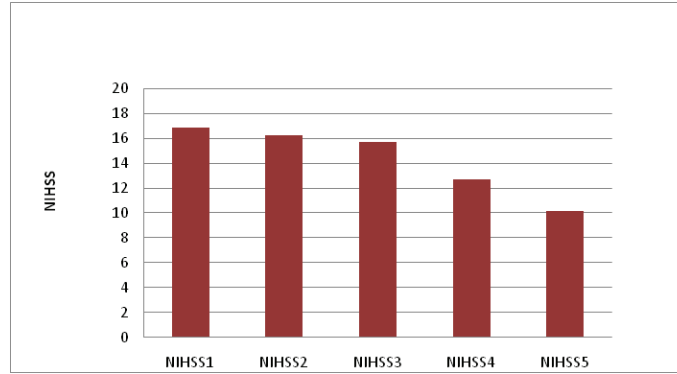
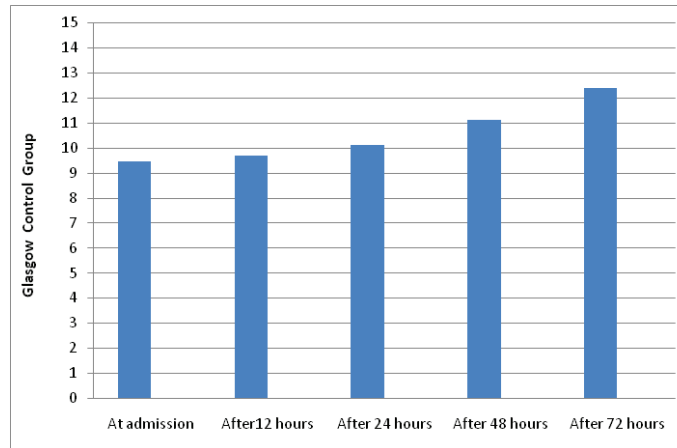


Fig. 7. Evolution of NIHSS in patients with ischemic Stroke (control Group)



NIHSS1: at admission; NIHSS2: after 24 hours; NIHSS3: after 48 hours; NIHSS4: after one week; NIHSS5: after 2 weeks.

Fig. 8. Glasgow Evolution in Control Group (Ischemic Stroke)



Statistical analysis

We decided compared in the analysis of the A2 group the Glasgow scale at the 12, 24, and 48 hours. We compared the groups separated by sex.

The analysis of Glasgow scale on samples cases and control were relevant statistically at the 12, 24, and 48 hours, with $(p=5.41 \times 10^{-5})$, $(p=5.21 \times 10^{-13})$, $(p=3.57 \times 10^{-17})$, respectively.

We decided compared in the analysis of the A2 group the NIHSS at the 24 hours, 48 hours, and one week. We compared the groups separated by sex.

The analysis of NIHSS on samples cases and control were relevant statistically at the 24 hours, 48 hours, and one week with $(p=1.00 \times 10^{-9})$, $(p=3.04 \times 10^{-19})$, $(p=3.67 \times 10^{-27})$, respectively.

We decided compared in the analysis of the A1 Group the Glasgow Scale at discharge. We compared the groups separated by sex.

The analysis of Glasgow Scale on samples cases and control were relevant statistically $(p=2.49 \times 10^{-6})$. (See Fig. 9 and 10)

Fig.9. Glasgow Comparison between Treated and Control Group in Stroke Patients.

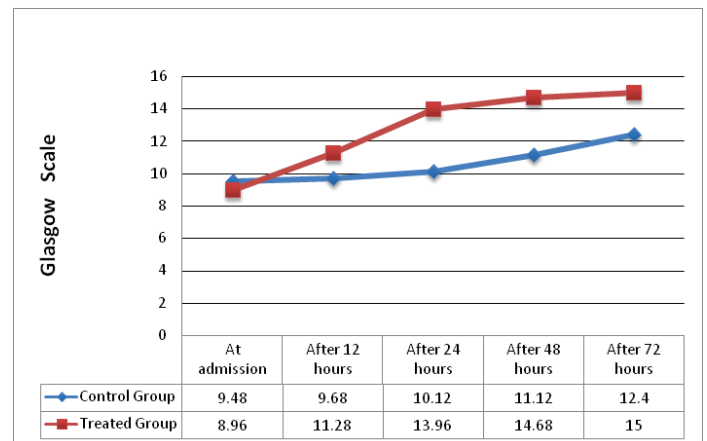
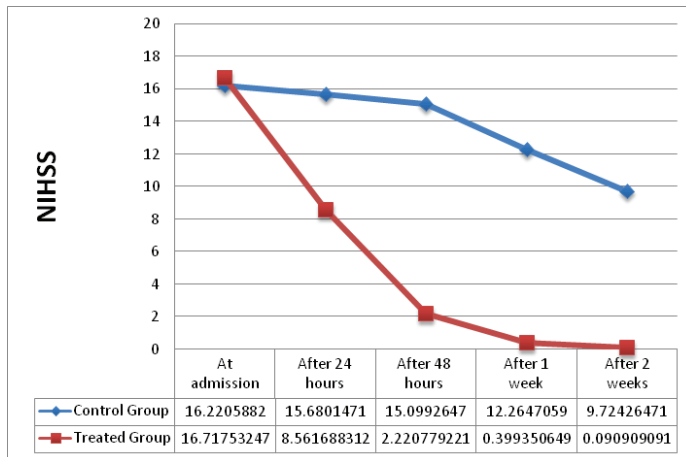


Fig.10. NIHSS Comparison Between Treated and Control Group in Stroke Patients



Discussion

A new neuroprotective treatment that combines low doses of an Inhibitor of glutamate release drug (Lamotrigine) with endothelial-protection drugs (L-Arginine +Tianeptine) seems to be a promising therapeutic approach. Our results show a novel and effective neuroprotective treatment. Until now there have been no previous studies that use neuroprotective treatment with L-Arginine and Tianeptine. This inedited study opens the door to the search of integral therapy. Our findings demonstrate that low doses of agents acting indirectly on glutamate release act better in stroke patients than usual doses of NMDA-Blockers as indicated in previous reports. (5,11,21,29) On the other hand our findings in patients with cerebral trauma confirm this assertion. To date, many cytoprotective drugs have reached the stage of pivotal phase 3 efficacy trials in acute stroke patients. Unfortunately, throughout the neuroprotective literature, the phrase "failure to demonstrate efficacy" prevails as a common thread among the many neutral or negative trials, despite the largely encouraging results encountered in preclinical studies. The reasons for this discrepancy are multiple and have been discussed (21), but recently, Papadia and Hardingham (2007) (40) explain it through the dichotomy of NMDA receptor signalling.

Failed Clinical Trials for Stroke with NMDA receptor antagonists

The calcium channels antagonist that has undergone the most extensive investigation in stroke is nimodipine (13,42). Several randomized controlled clinical studies have

conclusively demonstrated the effectiveness of nimodipine in preventing ischemic neurologic deficit and poor outcome secondary to aneurismal subarachnoid hemorrhage. (42) The most recent and extensive meta-analysis of 22 calcium antagonist trials, studying over 6,800 patients, failed to demonstrate any beneficial effect of treatment, a finding attributed to hypotension induced by both oral and intravenous administration of the drug. The lack of effect or presence of detrimental effect, of calcium antagonists may be due to the hypotension caused by blocking the vascular smooth muscle cells. Another plausible explanation is that the P13K-Akt pathway (phosphoinositide-3-kinase-Akt kinase cascade) is a key-signalling pathway responsible for pro-survival effects of NMDA-receptors activity that can be activated in a calcium-dependent manner. (40,18). On the other hand, the CREB (cAMP response element-binding protein) a key mediator of activity-dependent gene expression is strongly induced by NMDA-receptor activity and calcium. (40)

Glutamate antagonists N-methyl-D-aspartate (NMDA) receptor antagonists were the first class of acute stroke therapeutic agents to proceed from development in the laboratory to testing in humans, employing modern principles of clinical trial design, most importantly relatively early treatment. The potential utility of NMDA antagonists in stroke was first recognized when it was observed that a hypoxic or ischemic insult results in elevation of brain levels of the excitatory neurotransmitter glutamate. The excitotoxic theory of ischemic brain injury implicates glutamate as a pivotal mediator of cell death via ligand-gated receptors (NMDA and AMPA receptors). The NMDA receptor is a complex ligand-gated ion channel that requires activation by glutamate and glycine, as well as concomitant membrane depolarization to overcome a voltage-dependent block by magnesium ions. Selfotel (CGS19755) is a competitive NMDA receptor antagonist that limits neuronal damage in animal stroke models (13,14). Selfotel was evaluated in a randomized, double-blind, placebo-controlled, ascending dose phase 2a study. Neuropsychiatric adverse experiences were common, dose-related symptoms included hallucinations, agitation, confusion, dysarthria, ataxia, delirium, paranoia, and somnolence. Patients presented mild adverse experiences with Selfotel 1.5 mg/kg; however, when the dose was increased to 2 mg/kg given once or twice, adverse experiences occurred in all patients (15). The non-competitive NMDA antagonist dextrorphan was also evaluated in a pilot study (2) As with Selfotel, adverse effects of dextrorphan occurred in a dose-dependent manner. The reasons for this discrepancy

are well explained by Papadia S. And Hardingham E. (2007) (40): “NMDA-receptors are essential mediators of synaptic plasticity and also mediate aspects of development and synaptic transmission. However, when excessively activated, NMDA-receptors cause cell death in many neuropathological scenarios. During an ischemic episode, extracellular glutamate builds up due to synaptic release and impaired/reversed uptake mechanisms resulting in overactivation of NMDA-receptors. The destructive effects of excessive NMDA-receptor activity are in contrast to the recent findings that survival of several neuronal types is dependent on physiological synaptic NMDA-receptor activity (16). Thus, responses of neurons to glutamate or NMDA follow a bell-shaped curve: both too much and too little NMDA-receptor activity is potentially harmful” (30)

Magnesium (Mg²⁺) is an ideal neuroprotectant based upon its diverse mechanisms of action, low cost, ease of administration, wide therapeutic index, good blood-brain barrier (BBB) permeability, and established safety profile. Mg²⁺ ions endogenously function as a physiologic voltage-dependent block of the NMDA receptor ion channel and inhibitor of ischemia-induced glutamate release (28). Preclinical models show that **magnesium** reduces infarct volume with a dose-response relationship demonstrated within even 6 hours after stroke (28,52). Several pilot studies have already demonstrated the safety and tolerability of intravenous Mg²⁺ in acute ischemic stroke patients (33,34), The FAST-MAG (Field Administration of Stroke Therapy-Magnesium) (44), pilot study was an open-label evaluation of the safety and feasibility of paramedic-initiated magnesium therapy to stroke patients identified in the field by the Los Angeles Prehospital Stroke Screen (LAPSS). Greater than two-thirds of patients had a good functional outcome. Probably the ion Magnesium showed to be more effective treatment because its action is less aggressive in blocking the NMDA-receptors, allowing the effects on NMDA-receptors pro-survival signalling.

Our study has been made to develop strategies inhibiting glutamate-induced damage while avoiding the toxicity profile of direct NMDA receptor antagonism. In fact, we used in this study low doses of Lamotrigine to prevent down-regulation of neural receptors and to avoid the complete disappearances of the glutamate in the synapses, in order to allow the physiological NMDA-receptor signalling. To increase the effect of the treatment we added two unusual endothelial protection drugs (Tianeptine and L-Arginine) to prevent cell injury by other mechanisms like hypoxia and low blood flow. If we combining neuroprotective agents

that together have high potency by targeting multiple pathways then the citoprotection could be effective. The anti-epileptic drug lamotrigine inhibits glutamate release and has shown beneficial effects in a rodent model of focal cerebral ischemia when administered immediately after ischemia (9,22,45); however, a 2-h delay of treatment produced no effect on infarct volume or neurological outcome in two models. (47) To our knowledge, only one clinical stroke trials of Lamotrigine have been performed but with doses that could block the physiological NMDA-receptor signalling (6).

Taken in their entirety, the data suggest that monotherapy targeting a single neurotransmitter function may not provide sufficient neuroprotection to offer clinically meaningful benefit and the doses used produced neuro-psychiatric adverse experiences that could be explained by “The dichotomy of NMDA receptor signalling”.

The Dichotomy of NMDA receptor signalling

The NMDA subtype of ionotropic glutamate receptors plays a Jekyll and Hyde role in the mammalian central nervous system (40). There is a lot of evidence that indicates the NMDA-receptor activated plays a dual role. If the stimulus is intense or too low then NMDA-receptor activity promotes cell death. (40) The classical bell-shaped curve model of the neuronal response to NMDA or glutamate contends that intermediate, physiological NMDA-receptor activity levels are necessary for neuroprotection. We infer that our results are explained by this novel concept. The very low doses of Lamotrigine that we used were able to reach physiological NMDA-receptor activity levels. On the other hand is possible that the Lamotrigine may have prevented the action of Glutamate and Aspartate in the extrasynaptic NMDA-receptors. A recent study involving genome-wide expression analysis has extended the understanding of synaptic vs. extrasynaptic signalling (54). While synaptic NMDA-receptors activated a number of pro-survival genes, extrasynaptic NMDA-receptors failed to do this, and activated expression of a gene *Cla1* that kills neurons.

Serotonin and Ischemia

Serotonin plays an important role in ischemia. In humans, serotonin is concentrated in platelets and is released when platelets aggregated. The f-5HT plasma levels increase in vascular thrombosis secondary to platelet aggregation. Thus, this neurotransmitter increases the

platelet aggregation and vasoconstriction (31, 43) that worsens the ischemia. It has been demonstrated in cerebral arterioles of rabbits (31, 43) and in coronary arteries. (43). On the other hand, increased circulating catecholamines are responsible for the lowered p-5HT and the increased f-5HT registered during stressful situations, which trigger platelet aggregation. (24, 25). We used Tianeptine in this study in order to reduce the plasma levels of f-5HT owing to the fact that this drug enhances the platelet uptake of serotonin.

Recent studies show that Tianeptine targets the phosphorylation-state of glutamate receptors at the CA3 c/a synapse. This novel signal transduction mechanism for Tianeptine may provide a mechanistic resolution for its neuroprotective properties (20). On the other hand local Tianeptine has found to inhibit the activity of nitric oxide synthase (NOS) in the hippocampus (49).

Nitric Oxide and Agmatine

Other strategies of neuroprotection attack later stages of the ischemic cascade. Neuronal nitric oxide (nNO) synthesis is induced by stimulation of glutamate receptors, and nNO in turn has a number of complex actions relevant to ischemia and cell injury. Endothelium-derived NO (eNO) causes vasodilatation beneficial to ischemic brain, but nNO generates oxygen free radicals toxic to cells. The usefulness of NO modulation in stroke likely will hinge on the ability to favourably manipulate the beneficial and deleterious effects of NO. We decided to use low doses of L-Arginine like a precursor of NO due to the important role that endothelial NO plays in the cerebral vasodilatation, vascular remodelling and angiogenesis in human and animal models (21,26). In our opinion low doses of L-Arginine increase NO at low levels that are not toxic to cells. On the other hand, L-Arginine blocks the release of somatostatine thus increases the Grow Hormone levels that have anti stress and cell-reparative properties (26). Moreover, L-Arginine is a precursor of the novel neuroprotective Agmatine that has been shown to be neuroprotective in trauma and ischemia models through regulation of endothelial nitric oxide synthase, reduction of brain edema and glutamate release (19,36). On the other hand, agmatine can reduce brain infarction through minimizing neuroinflammation and can lessen the danger of post-stroke infection from depression of the immune system after stroke (48).

Free radicals production occurs during ischemia and reperfusion and contributes to the neuronal injury after stroke. In order to avoid the effect of NO as a free radical

we reduced the doses of L-Arginine after two weeks of treatment.

Our study has potential limitations. One of them is the imbalance in the two sample groups. The A1 group included cerebral trauma patients with ostensible cerebral damage. The A2 group included only cerebral stroke patients with only ischemic damage. In the former the most important evaluation was the use of low doses of lamotrigine to avoid the progression of the cerebral injury (Low doses owing to the fact that the experience in neuroprotection studies with NMDA blockers shows many side effects). In the latter the key was the use of three different mechanisms to avoid the progression and damage of the penumbra zone. Other limitation was the fact that the A1 group began the treatment 4 or more days after the cerebral trauma occurred, and this group included only patients catalogued as "Did not respond to the usual therapy". There is a disadvantage because there is a lot of information concerning the beneficial use of early neuroprotective treatment in phase I and II studies. Other limitation is that we began the treatment in the A2 group earlier and we cannot compare it to the A1 group.

Our results show that the use of Lamotrigine + L-Arginine + Tianeptine in cerebral trauma and ischemic stroke results in higher survival rates and lower rates of neurologic complications at one week since the onset of the symptoms. On the other hand, this treatment significantly reduces the clinical-recuperation-time of Glasgow scale and NIH stroke scale when it was compared with the conventional treatment.

Since the Food and Drug Administration has recognized that stroke is a serious and life-threatening condition, making it eligible for accelerated approval, we believe that this treatment can be extrapolated to institutions with resources in stroke and cerebral trauma trial. An inexpensive, effective and safe neuroprotective treatment could be evaluated in a large-scale, multicenter, double-blind, placebo-controlled, randomized trials.

Finally, we believe that delineating the mechanism underlying the vulnerability of the central nervous system to diverse insults should lead to new therapeutic interventions that affect the outcome positively.

Acknowledgements: *We are indebted to the teacher Reygar Bernal and Miss Deika Terant for there help in the elaboration of the manuscript.*

We are grateful to Dr. Ignacio S. Torres-Alvarado for statistical support.

References

- 1) A report of the American Heart Association 2011. Heart Disease and stroke statistical 2011 update (2011). *Circulation*. (123): e18-e209
- 2) Albers G, Atkinson R, Kelley R (1995) Safety, tolerability and pharmacokinetics of the N-methyl-D-aspartate antagonist dextrorphan in patients with acute stroke. *Stroke*; 26:254-258.
- 3) Asplund K, (1996) Stroke in Europe: widening gap between East and West. *Cerebrovasc Dis*. (6): 3-6.
- 4) Bamford J, Sandercock P, Dennis M, Warlow C, Burn J. (1991) Classification and natural history of clinically identifiable subtypes of cerebral infarction. *The Lancet*; 337: 1521-1526.
- 5) Brott T, Bogousslavsky J. (2000) Drug: Treatment of Acute Ischemic Stroke. *N Engl J Med*; (343):710-722.
- 6) Chen S, Lee J, Yang, et al. (2002) Combination Therapy for Ischemic Stroke: Potential of Neuroprotectants Plus Thrombolytics. *Am J Cardiovasc Drugs*; 2: 303-13
- 7) Choi D. (1988) Manipulating Glutamatergic Neurotransmission for therapeutic gains: The example of brain ischemia. Speaker, NIH National meeting. *The Glutamate Cascade: Common Pathways of Central Nervous System Disease States*.
- 8) Christopher J, Markkuk K, Kennedy R. (2004) Targeting Neuroprotection Clinical Trials to Ischemic Stroke Patients With Potential to Benefit from Therapy. *Stroke*; 35:2111.
- 9) Crumrine R, Bengstrand K, Cooper A. (1997) Lamotrigine Protects Hippocampal CA1 Neurons from Ischemic Damage After Cardiac Arrest. *Stroke*; 28: 2230-7
- 10) Cui X, Chopp M, Zacharox A, Zhang C, Roberts C, Chen J. (2009) Role of endothelial Nitric Oxide Synthetase in arteriogenesis after stroke in mice. *Neuroscience*; 159 (2): 744-750
- 11) Devuyst G, Bogousslavsky (1999) J. Clinical trial update: Neuroprotection against acute ischemic stroke. *Current Opin Neurol*. (12): 73-9
- 12) Graham S, Chen J, Sharp F. (1993) Limiting Ischemic Injury by Inhibition of excitatory Amino Acid Release. *J Cereb Blood Flow Metab*; 13: 88-97.
- 13) Grotta J. (1991) Clinical aspects of the use of calcium antagonists in cerebrovascular disease. *Clin Neuropharmacol*; 14:373-390.
- 14) Grotta J, Picone C, Ostrow P. (1990) CGS-19755, a competitive NMDA receptor antagonist, reduces calcium-calmodulin binding and improves outcome after global cerebral ischemia. *Ann Neurol*; 27:612-619.
- 15) Grotta J, Clark W, Coull B. (1995) Safety and tolerability of the glutamate antagonist CGS 19755 (Selfotel) in patients with acute ischemic stroke: results of a phase IIa randomized trial. *Stroke*; (26): 602-605.
- 16) Hardingham G, Bading H. (2003) The Yin and Yang of NMDA Receptor Signalling. *Trends Neurosci*; 26(2): 81-9
- 17) Ikonomidou C, Turki L, (2002) Why did NMDA receptor Antagonists Fail Clinical Trials for Stroke and Traumatic Brain Injury? *The Lancet Neurology*; (1) 383-386.
- 18) Joyal J, Burks D, Pons S, Matter W, Vlahos C, White M. (1997) Calmodulin Activates Phosphatidylinositol 3-Kinase. *J Biol Chem*; 272(45): 28: 183-6
- 19) Kim J, Lee Y, Park K, Lee W, Lee J. (2010) Agmatine attenuates brain edema through reducing the expression of aquaporin-1 after cerebral ischemia. *J Cereb Blood Flow Metab*; 30(5): 943-9
- 20) Kole M, Swan L, Fuchs E. (2002) The antidepressant Tianeptine persistently modulates glutamate receptors currents of the hippocampal CA3 commissural-associational synapse in chronically stressed rats. *Eur J Neurosci*; 16(5):807-816.
- 21) Labiche L, Grotta J. (2004) Clinical Trials for Cytoprotection in Stroke. *NeuroRx*; 1:46-70.
- 22) Leach M, Swan J, Eisenthal D, Dopson M, Nobbs M. (1993) BW619C89, a glutamate release inhibitor, protects against focal cerebral ischemic damage. *Stroke*; 24:1063-1067.
- 23) Lechin F, Van Der Dijs B, Orozco B. (2004) Elective setting, platelet serotonin and thrombotic events. *Platelets*; 15(7): 462
- 24) Lechin F, Van Der Dijs B. (2001) Plasma serotonin, bronchial asthma and tianeptine. *Clin Physiol*; 21(6): 723-4
- 25) Lechin F, Van Der Dijs B. (2003) Tianeptine, plasma serotonin and pulmonary hypertension. *Lancet*; 361:87
- 26) Lechin F, Van der Dijs B, Lechin M. (2002) Stress versus depression In: Lechin F, Van der Dijs B. *Neurocircuitry and neuroautonomic disorders. Reviews and therapeutic strategies*. Karger, Basel; 52-54.
- 27) Lipton S, Rosenberg R. (1994) Mechanisms of disease: excitatory amino acids as a final common pathway in neurologic disorders. *N Engl J Med*; 330:613-622.
- 28) Lipton S, Nicotera P. (1998) Calcium, free radicals and excitotoxins in neuronal apoptosis. *Cell Calcium*; (23): 165-171.
- 29) Lipton S. (2004) Failures and Successes of NMDA Receptor Antagonists: Molecular Basis for the Use of

- Open-Channel Blocker like Memantine in the treatment of Acute and Chronic Neurologic Insults. *NeuroRx* 2004; 1:101-110.
- 30) Lipton J, Nakanishi N. (1999) Shakespeare in Love with NMDA Receptors? *Nature Medicine*; 5(3): 270-271
- 31) Marshall J, Kontos H. (1990) Endothelium-Derived Relaxing Factors A Perspective From In Vivo Data. *Hypertension*; 16:371-386.
- 32) Muir K. (2002) Magnesium in stroke treatment. *Postgrad Med J*; 78:641-645.
- 33) Muir K, Lees D. (1995) A randomized, double-blind, placebo-controlled pilot trial of intravenous magnesium sulphate in acute stroke. *Stroke*; 126:1183-88.
- 34) Muir K, Lees D. (1998) Dose optimization of intravenous magnesium sulfate after acute stroke. *Stroke*; 29:918-923.
- 35) Muir K, Holzapfel L, Lees K. (2000) Phase II clinical trial of sipatrigine (619C89) by continuous infusion in acute stroke. *Cerebrovasc Dis*; 10:431-436.
- 36) Mun C, Lee W, Park K, Lee J. (2010) Regulation of endothelial nitric oxide synthase by agmatine after transient global cerebral ischemia in rat brain. *Anat Cell Biol*; 43(3):230-40
- 37) Neuroprotection as initial therapy in acute stroke: Third report of an Ad Hoc Consensus Group Meeting: the European Ad Hoc Consensus Group. (1998) *Cerebrovasc Dis* 8:59-72.
- 38) Palmer R, Ferrige A, Moncada S. (1987) Nitric Oxide release accounts for the biological activity of endothelium-derived relaxing factor. *Nature*; 327:524-26
- 39) Palmer R, Rees D, Ashton D. (1988) L-Arginine is the physiological precursor for the formation of nitric oxide in endothelium-dependent relaxation. *Biochem Biophys Res Commun*; 153:1251-1256.
- 40) Papadia S, Hardingham E, (2007) The Dichotomy of NMDA receptor Signalling. *Neuroscientist*; 13(6): 572-579
- 41) Papapetropoulos A, Garcia-Cardeña G, Madri J, Sessa W. (1997) Nitric Oxide Production contributes to the angiogenic properties of vascular endothelial growth factor in human endothelial cells. *J Clin Invest*; 100 (12): 3131-9
- 42) Petruk K, West M, Mohr G. (1988) Nimodipine treatment in poor-grade aneurys patients: Results of a multicenter double-blind placebo-controlled trial. *J Neurosurg*; 68:505-517.
- 43) Sanchez C, Marin J. (1990) Endothelium-Derived Contractile Factors. Review. *Clin Pharmac*; 31(5):589-603.
- 44) Saver J, Kidwell C, Leary M. (2001) The field administration of stroke therapy-magnesium (FAST-MAG) pilot trial. Abstract presented at the ongoing clinical trials session, 26th International Stroke Conference, Fort Lauderdale, FL, Feb 2001.
- 45) Smith S, Meldrum B. (1995) Cerebroprotective effect of Lamotrigine after focal ischemia in rats. *Stroke*; 26:117-121.
- 46) Stuehr, D. (1999) Mammalian nitric oxide synthases. *Biochim Biophys Acta*; 1411:217-230
- 47) Traystman R, Klaus J, DeVries A. (2001) Anticonvulsant lamotrigine administered on reperfusion fails to improve experimental stroke outcomes. *Stroke*. 2001; 32:783-787.
- 48) Uranchimeg D, Kim J, Kim E J, Lee J. (2010) Recovered changes in the spleen by Agmatine treatment after transient cerebral ischemia. *Anat Cell Biol*; 43(1): 44-53.
- 49) Wegener G, Volke V, Harvey BH, Rosenberg R. (2003) Local, but not systemic, administration of serotonergic antidepressant decreases Hippocampal nitric oxide synthase activity. *Brain Res*; 959: 128-34
- 50) Wiard R, Dickerson M, Beek O. (1995) Neuroprotective Properties of the novel Antiepileptic Lamotrigine in a Gerbil Model of Global Cerebral Ischemia. *Stroke*; 26: 466-72
- 51) Xion Z, Zhu X, Chu X. (2004) Neuroprotection in Ischemia: Blocking calcium-permeable acid-sensing ion channels. *Cell*; 118:697-98.
- 52) Yang Y, Li Q. (2000) Survival and histological evaluation of therapeutic window of post-ischemia treatment with magnesium sulphate in embolic stroke model of rat. *Neurosci Lett*; 285:119-122.
- 53) Yong-Hong Y, Wen-Chao G, Wei-Wen S, Tao S, et al. (2008) Neuroprotection of Lamotrigine on Hypoxic-Ischemic Brain Damage in Neonatal Rats: Relations to Administration Time and Doses. *Biologics*; (2): 339-344
- 54) Zhang S, Steijaert M, Lau D, Schulz G. Et al. (2007) Decoding of Genomic Programs Specifying Neuronal Survival and Death. *Neuron*; 53(4): 549-62

Metastatic Breast Cancer: An Evaluation of the Safety and Effectiveness of High Dose Chemotherapy with Autologous Stem Cell Transplant

Martin F. Taylor (MSc)

Master of Science and candidate to PhD in Science at the Department
of Science of the School of Doctoral Studies of the EU

[Email: mf.taylor@dsc.sds.eu](mailto:mf.taylor@dsc.sds.eu)

Abstract

Breast cancer is the leading cause of death for American women. Metastatic breast cancer, the focus of this paper, is defined as cancer that originated in the breast and has moved to lymph nodes and other bodily tissues. The treatment researched here is high-dose chemotherapy (HDC) followed by an autologous stem cell transplantation (ASCT). The safety and efficacy of HDC/ASCT are still under examination. Some arguments state that the effectiveness of HDC/ASCT may vary according to these non-treatment related factors: age, stage of disease at diagnosis, hormone receptor status, use of prior adjuvant chemotherapy, initial disease-free interval, response to prior chemotherapy, and number and sites of metastases. At this time it is still unclear if HDC/ASCT is a more effective treatment than traditional chemotherapy, as the survival rate is not significantly higher, while toxicity is higher. The health and safety of the women is important, and a safe treatment is one that does not kill the patient and does not result in permanent side effects. It has been concluded that higher levels of chemotherapy induce higher levels of toxicity and mental disturbances not limited to the following: nausea, myelosuppression, infection, mucositis and diarrhea, depression, stress, and fatigue. There is a low mortality rate associated with the actual transplant itself. The practice of re-infusion of stem cells has become a valuable method of hematopoietic rescue, as longer survival and few long term illnesses are associated with it. In the future more research should go into identifying the specifics that will define the women who will benefit from HDC/ASCT. The development of drugs that would not harm non-cancerous cells would be the ideal solution to cancer.

Breast cancer is the leading cause of death for American women. In 2002, over 203,000 American women suffered from the disease and nearly 40,000 died (Williams, 2002). Surgical treatment to directly remove cancerous tissue involves either a lumpectomy or mastectomy where regional lymph node tissue is often removed to assess for cancer spread. Conventional treatments used in addition to surgery can include radiation therapy, hormone treatment, and standard-dose chemotherapy (Dervan, 2001). In some unfortunate cases, the cancer spreads beyond the regional lymph nodes and then disperses to other parts of the body; this is known as metastatic breast cancer (MBC) (Prucha and Bellenir, 2001). The focus of this paper is on the treatment of MBC, the most harmful and least curable form of breast cancer.

High dose chemotherapy and autologous stem cell transplantation is another method of treatment that has recently become available to women with more advanced

forms of MBC. Since stem cells are undifferentiated, they are a useful form of rescue in that they naturally replace depleted cells in the body. The use of stem cells as a rescue treatment for cancer and immuno-recovery was first practiced in the treatment of leukemia. After many years of trials on animals as well as humans, and confidence levels in the procedure were heightened, stem cell rescue was thought to be an ideal method for rebuilding the immune system. Stem cell transplants have since been used in addition to HDC to compensate for the severe damage that the immune system incurs in the process (Thomas *et al.*, 1999).

Often, the first step in this treatment program is a low dose chemotherapy treatment tailored to eliminate the cancer and induce remission (induction chemotherapy). Stem cells are then collected from the patient's own bone marrow and are frozen for later use. HDC/ASCT differs from traditional chemotherapy (TC) in that stem cells

are not delivered to the patient in TC. Also, TC delivers chemicals to the patient in lower doses and over longer periods of time, often months and sometimes years. HDC differs in that the administration of the chemicals is given in a high dose, and for three to four days at a time. This treatment is generally repeated approximately two weeks later, and the entire cycle may be repeated several times. Thus, the HDC process can have a much stronger effect in only a fraction of the time necessary to complete TC. Once HDC is completed, the stem cells are infused back into the body to rebuild the immune system and recover lost hematopoietic cells (Thomas *et al.*, 1999). However, the procedure can be taxing on the body and has both positive and negative effects. It is important to substantiate whether or not HDC/ASCT is a viable treatment option, and for this reason uses of such treatment are persistently examined. This paper will survey available literature to determine the true safety and efficacy of this procedure as a treatment for breast cancer.

Current Research

Efficacy

For the past decade, high-dose chemotherapy with autologous stem cell transplantation has become increasingly popular for the treatment of women with breast cancer. Generally, it is agreed that higher dose intensity has a direct correlation to higher response rates; however, these positive response rates often last for only a short interval (Williams, 2002). According to numerous studies, approximately 10%-20% of HDC/ASCT patients are able to attain long-term progression-free survival (measured from time of the first transplant until either progression of the disease or death) following treatment (Schneeweiss *et al.*, 2002). Bias in patient selection is important to consider and it is possible that some patients could experience similar results through traditional chemotherapy, which would renounce any reason to go through the extra strain of HDC.

Some researchers declare that the effectiveness of HDC/ASCT may vary according to non-treatment related factors such as patient age, hormone receptor status (HER2-neu, estrogen, and progesterone), location and quantity of metastatic sites, and the disease stage at diagnosis. Several other treatment-related variables may also influence the efficacy of HDT/ASCT, including use of prior adjuvant

chemotherapy, previous disease-free intervals, and response to prior chemotherapy. Each of these factors needs to be taken into account when judging the effectiveness of the treatment.

Most studies seem to agree that younger patients (<45-60 years) have a better chance of survival mainly due to stronger tolerance of such intense treatment (Winter, 1997). Also, some researchers questioned whether or not premenopausal patients may have better results with HDC/ASCT than post-menopausal patients, but no correlation has been found (Schneeweiss *et al.* 2002).

The status of various hormone receptors in cancerous tissue is another prognostic factor. Overexpression of human epidermal growth factor receptor (HER2/neu) occurs in 25-30% of metastatic breast cancer (MBC) patients, and is generally associated with a curtailed survival time (Hensel, *et al.*, 2002). However, it is hypothesized that HER2/neu overexpression might confer an improved response to various chemotherapy drugs. In 2001, Kim *et al.* conducted a study that evaluated the effects of HER2/neu on the outcome of HDC/ASCT treatment. They treated a group of MBC that showed overexpression of HER2/neu and compared the results with another group given the same treatment whose cancers did not demonstrate such overexpression. Only 21% of patients with HER2/neu overexpression achieved a complete response following HDC/ASCT while a complete response was attained by 42% of patients whose cancers that lacked this characteristic. It was concluded that the presence of HER2/neu overexpression does not enhance the effectiveness of HDC. However, Kim *et al.* also found that 12 of the 19 patients with HER2/neu overexpression had an absence of estrogen receptor (ER) and progesterone receptor (PR) expression, while 29 of 40 patients who had HER2/neu-negative tumors showed a positive receptor status for ER and PR. Overall, 69% of their patients (41 of 59), showed a correlation between ER/PR status and expression of HER2/neu. This is interesting because results from a study conducted by Rowlings *et al.* in 1999 (to be discussed later in this paper) showed that those with a positive ER and PR status experienced much higher success rates during HDC/ASCT than those with negative hormone receptor status. In 2002, Schneeweiss *et al.* further supported this finding by demonstrating that patients with this positive status were more likely to achieve long-term progression-free survival following HDC/ASCT. Given these three studies, it may be that HER2/neu is not an independent prognostic factor that

influences HDC/ASCT outcome, but rather predicts ER and PR status which directly correlates with the effectiveness of such treatment. In other words, overexpression of HER2/neu often predicts the absence of ER and PR, which has poor prognostic effects on MBC treated with HDC/ASCT.

In 1999, Rowlings *et al.* conducted a study specifically to determine the effectiveness of HDC/ASCT in relation to both treatment-related and non-related variables. Treatment was administered to and effectiveness was analyzed for 1,188 women with either MBC or advanced breast cancer (where cancer may return to the breast and/or regional lymph nodes after a disease-free interval) (Rowlings *et al.* 1999). For our purposes we will focus our discussion on the MBC patients. Women tested were between the ages of 18-70, and had different characteristics with regards to hormone receptor status, disease-free interval, sites of disease, chemotherapy responsiveness. Rowlings *et al.* found several characteristics to be statistically significant for treatment outcome: stage of disease at diagnosis, hormone receptor status (as mentioned previously), use of prior adjuvant chemotherapy, initial disease-free interval, response to prior chemotherapy, and number and sites of metastases.

Women who reached progression to metastases by the time of diagnosis had a higher risk of treatment failure than those who were diagnosed at an earlier stage of the disease. Findings also demonstrated that women who had not received prior adjuvant chemotherapy were likely to experience treatment failure. This point is unclear because the authors fail to specify whether they mean prior adjuvant therapy as induction therapy to HDC or as an earlier attempt to treat the disease. If their intent was the latter, we may interpret this information by concluding that the reason Rowlings *et al.* found that metastases at time of diagnosis had poor prognostic effects on HDC/ASCT is because these women, being at a severe stage of disease, less likely received any prior treatment. However, most other cases argue that prior therapy would confer a negative effect on results of HDC/ASCT due to a possibility that the cancer could develop resistance to such chemotherapy treatment (Williams, 2002). Rowlings *et al.* also found women with a disease-free interval less than 18 months frequently experience treatment failure when compared to women who achieved a disease-free interval of greater than 18 months. Another more recent study claims that length of disease-free interval only significantly increases the success of HDC/ASCT if it is longer than 24 months (Schneeweiss *et al.* 2002). In either case, it is possible that women may

not need to receive prior chemotherapy, but perhaps just some form of treatment, which allows them to achieve this disease-free interval. Also found in accordance with other studies, is that women who attain complete response or partial response to prior chemotherapy treatments will have a more positive prognosis following HDC/ASCT treatment (Rowlings *et al.*, 1999).

Finally, Rowlings *et al.* found that women with metastases to the liver, central nervous system, or three or more sites in other organs had a very poor probability of post-treatment progression-free survival. Women who had less than three sites of metastases had a better prognosis and chance for progression-free survival. In a study conducted by Schneeweiss *et al.* in 2002, 81% of patients who achieve long-term progression-free survival and 76% of patients who attained a complete response to HDC/ASCT had only one metastatic site. Both studies agreed that women who had metastases in liver or central nervous system would most likely experience failure of HDC/ASCT. Overall, Rowlings *et al.* outlined several risk factors that demonstrate that HDC/ASCT is not a procedure that all women should consider unless their cancers possess the favorable prognostic factors previously detailed.

Other factors that are particularly important when evaluating the effectiveness of HDC/ASCT are those pertaining specifically to the procedure. Such factors include dose levels of chemotherapy, number of treatment cycles dose of infused stem cells.

In 2002, Schneeweiss *et al.* conducted a study in order to determine the influence of chemotherapy dose and number of cycles on effectiveness on HDC/ASCT. One hundred twelve patients received induction therapy following be either one, two, or three cycles of HDC/ASCT of varying intensities ranging from 25.44 to 81.18 total units (mg/m²/week) (see Table 1).

Table 1. Dose intensity (in units) of chemotherapy regimens and number of patients who received those treatments.

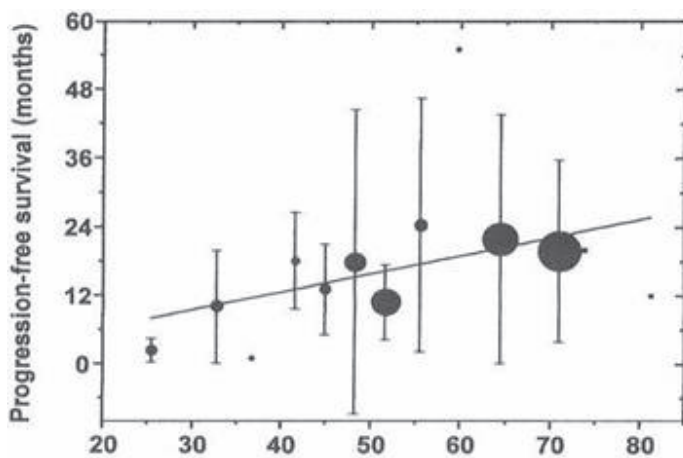
Intensity _{IND}	Intensity _{HD}	Intensity _{TOT}	No. of patients
One cycle of HDC (n=29)			
0	25.44	25.44	3
9.99	22.74	32.73	4
13.95	22.74	36.69	1
18.84	22.74	41.58	2
25.20	25.44	50.64	2
25.20	26.58	51.78	17

Intensity _{IND}	Intensity _{HD}	Intensity _{TOT}	No. of patients
Two cycles of HDC (n=48)			
0	44.94	44.94	3
18.84	29.16	48.00	5
9.99	38.46	48.45	5
9.99	45.48	55.47	4
14.14	45.48	59.62	1
18.84	45.48	64.32	17
19.41	45.48	64.89	2
19.98	45.48	65.46	1
25.20	44.94	70.14	9
28.26	45.48	73.74	1
Three cycles of HDC (n=41)			
9.99	54.18	64.17	7
9.99	61.20	71.19	33
19.98	61.20	81.18	1

Intensity_{IND}, dose intensity of induction chemotherapy; Intensity_{HD}, intensity of high dose chemotherapy; Intensity_{TOT}, dose intensity of totally chemotherapy (Intensity_{IND} + Intensity_{HD}), 1 unit = 1 mg/m²/week (Schneeweiss *et al.*, 2002)

Schneeweiss *et al.* agreed that there was a definite relationship between dose and survival time and found that an increase of ten units in intensity demonstrated an increase in progression-free survival time by an average of three months (see Figure 1).

Figure 1. "Progression-free survival (mean ± standard deviation) of different metastatic breast cancer patient subgroups who received at least one cycle of high-dose chemotherapy compared with the total dose intensity of the chemotherapy those patients received.



Each circle represents a subgroup of all 118 patients who received chemotherapy with almost identical dose intensity totals. Circle size is proportional to the number of patients" (Schneeweiss *et al.*, 2002).

Specifically, they found that patients with a previous progression-free survival time of greater than 24 months after prior adjuvant chemotherapy had the greatest response to intensities greater than 55 units. The authors did not conclude that the manner by which total intensity was divided (whether by one, two, or three cycles of HDC/ASCT) was relevant to the outcome of HDC/ASCT. Unfortunately, there are not many current studies that analyze this same aspect of HDC/ASCT, and the results from this study must be cautiously interpreted given the small sample of patients.

The dose of stem cells reinfused into the patient may also affect the success of the HDC/ASCT treatment. Hensel *et al.* conducted an experiment that analyzed success rates in 120 patients treated with HDC and then infused with varying doses of stem cells ranging from 2.3 x 10⁶ to 49.8 x 10⁶ stem cells per kilogram of body weight (median 7.8 x 10⁶) (Hensel *et al.* 2002). In their analysis of treatment outcomes, Hensel *et al.* concluded that a lower amount of stem cells per kilogram of body weight correlates with a substantially reduced overall survival time and that a higher stem cell dose frequently better the outcome. However, again, there has not been sufficient research looking into this aspect of the treatment, and further research with larger samples is necessary in order to make reliable conclusions.

Another topic in the debate over the efficacy is that the chance of failure in HDC/ASCT may be as great as or even greater than that of traditional chemotherapy (TC). While treatment of MBC with HDC/ASCT has consistently shown greater complete response rates (70-100%) than has TC (60-70%) (Folet *et al.*, 1999 and Williams, 2002), the data given by Schwartzberg *et al.* indicates that there is not a significantly greater number of survivors that underwent HDC/ASCT (1999). That said, it is unclear if HDC/ASCT is a better choice. Results from a more recent study conducted by Berry *et al.* seem to substantiate the findings Schwartzberg *et al.* The study by Berry *et al.* compared efficacy of TC versus that of HDC/ASCT. Patients were divided into two groups: those treated with TC and those treated with HDC/ASCT. The study explicitly mentions a number of biases and methods used to compensate for them. First, preferred candidates for ASCT are those who have experienced a complete response or partial response to induction therapy with TC prior to any exposure to HDC. To eliminate this bias, Berry *et al.* restricted their study to patients who displayed a chemosensitive disease and who had previously achieved complete response or partial response before protocol therapy for the TC cohort and before HDC for the HDC/ASCT cohort. Second, since HDC/ASCT

patients consisted only of women who had previously been untreated for MBC, a bias could have arisen from the trend that HDC is frequently given as an alternative to women whose disease is non-responsive to other forms of treatment. To compensate for this, Berry *et al.* excluded from the TC cohort women who had received even one treatment for MBC. Yet another factor creating possible bias is that HDC is more likely to be offered to “younger women, women with good performance status, and women who are at high risk of experiencing unsuccessful conventional chemotherapy” (Berry *et al.*, 2002).

The above discrepancy were adjusted for by evaluation of the following variable for each cohort: “patient age, performance status, sites of metastases, disease free interval, hormone receptor status, and whether the patient had received prior adjuvant chemotherapy” (Berry *et al.*, 2002). Exclusion criteria included age greater than 65, insensitivity to chemotherapy, lack of response

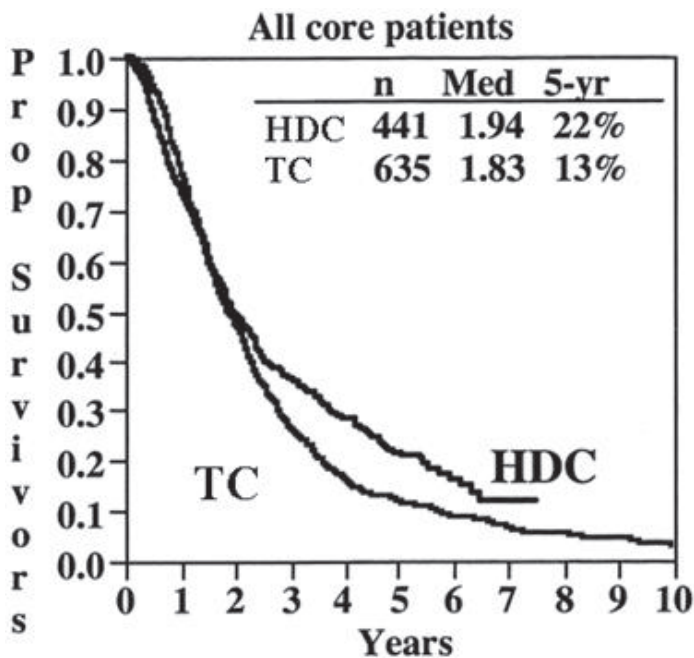
information, or this treatment being a second- or third-line therapy. Ultimately, 874 of 1,509 from the TC cohort and 747 of 1,188 from the HDC cohort were eliminated from the studied groups (see Table 2).

Table 2. Chronologic Application of Exclusion Criteria to Patients in HDC and TC Cohorts

Characteristic	TC	HDC
No. patients prior to exclusion	1,509	1,188
Age > 65	313	1
Not sensitive to chemotherapy	456	498
No response Total dose intensity (mg/m2/week)		
Second- or third-line therapy	0	197
Total excluded	874	747
Total included	635	441

TC, traditional chemotherapy; HDC, high dose chemotherapy (Berry *et al.*, 2002)

Figure 2. Comparison of TDC with HDC for core patients



Area between survival curves signifies the estimated advantage of life expectancy for HDC. (Approx. 4 mos.)
 HDc, high dose chemotherapy; TC, traditional chemotherapy; n = number of patients; Med, median survival time; 5-yr, percent of patients surviving after 5 years.

Figure 2 (Berry *et al.*, 2002) displays the survival of patients in each of the two cohorts. During the initial fifteen months following treatment survival for the TC cohort is higher (See Table 3).

Table 3. Statistics on Mortality and Survival Rates

		HDC	TC
100-date	mortality rate	4.5%	0.8%
3-year	survival probability	37%	27%
5-year	survival probability	22%	13%

Median survival times for TC and HDC/ASCT were very similar: 1.83 and 1.94 years, respectively (Berry *et al.* 2002). Moreover, Berry *et al.* also clarify that any increased survival linked to HDC/ASCT was often not notable until three years later.

Very few studies showing that HDC/ASCT has an advantage over TC are available. In 2000, the Dutch National Study “showed a modest survival benefit from high-dose chemotherapy” when compared with TC (Schrama *et al.* 2002), but again this analysis was conducted with a small sample of 284 patients and the results have not been replicated by further studies. In 2001, the American Intergroup study found that relapse was less likely to occur in patients who underwent HDC/ASCT (Schrama *et al.* 2002), but as is the case with other studies that found a marginal advantage, the small benefit was counterbalanced by frequent toxicity.

Furthermore, any benefits are most likely limited to a select number of candidates who have the favorable prognostic characteristics discussed earlier. A clear subgroup of candidates who will definitely benefit from the

intense HDC/ASCT procedure has yet to be identified. As with any newly developed treatment, failure rates among those treated is a serious concern. At this point in time the available data indicates that the benefits of HDC/ASCT do not significantly outweigh the risk of failure or potential toxicity.

Safety

As above, in addition to examining the statistics for effectiveness of HDC/ASCT, safety and potential side effects are also important. Some may consider a safe treatment to be one that does not ultimately kill the patient. In addition to fatalities, mental and physical traumas also contribute to the overall success or failure of the treatment. The most troubling aspects of treatments are the physical and emotional distresses that the women must endure in the hopes that the cancer will be eliminated. A safe treatment is defined as one that will not cause long-term harm to the woman and for which side effects are limited or short-lived.

High dose chemotherapy and stem cell transplantation is a treatment modality that varies between doctors, patients, and studies. There are several common combinations of drugs: cyclophosphamide and thiotepa, with or without carboplatin; melphalan, etoposide and carboplatin; and carmustine (BCNU), cisplatin, and cyclophosphamide (Cossaart *et al.*, 2003). The drugs are given intravenously over differing periods of time. The study by Genre *et al.* showed that recoverable side effects can still be maintained at dose intensities increased up to 1.25 times for doxorubicin and five times for cyclophosphamide. The authors used various time intervals and levels of chemotherapy, with the stem cells being collected on the first and thirteenth day of chemotherapy. The stem cells were reinfused into the patient after the third and fourth cycle of chemotherapy, and again on the third day after the administration of chemotherapy was complete. It was concluded that the level of intensity can safely be increased (to a maximal point) with no increase in toxic death and has a statistically similar survival rate. It was observed that the higher intensity treatments did produce higher levels of toxicity (Genre *et al.*, 2002). The purpose of this study was to achieve the highest possible efficacy by using the most tolerable amount of drugs. Unfortunately, the results did not offer an optimal dose. Other studies have also shown that while all HDC has unpleasant side effects, there is no outstanding result from higher strength doses, as the efficacy is not dramatically improved and toxicity is increased.

The human body is vulnerable to chemotherapy. The drugs are specifically designed to kill cancerous cells, but healthy cells are also destroyed in the process. This cell depletion is more pronounced with high doses of chemotherapy than with traditional chemotherapy, which is why stem cell transplants were proposed as an accompaniment for the treatment. Most of the healthy cells that are affected are part of the hematopoietic (blood) system, as not unlike cancerous cells, these cells reproduce rapidly. The HDC almost entirely eliminates the immune system and leaves the body susceptible to infections, and could result in a slow recovery. This high rate of devastation is what fuels the need for the transplant (Byers, 2000). There is a low mortality rate and low toxicity associated with the actual transplant because the source cells come from the patient undergoing treatment (autologous). Therefore, there will be a negligible chance for rejection from the body. Again, Schwartzberg *et al.* noted that the efficacy of this treatment was not significantly greater than TC. Although, it was also observed that platelets and neutrophils were recovered much more rapidly with the stem cell transplant than if they were left to recover without it. The importance of this is seen in the hematopoietic recovery levels. Patients who receive the transplants have more energy and have a more stable immune system following the procedure due to the increased number of cells replicating in the hematopoietic system as well as throughout the body (Schwartzberg *et al.*, 1999). This observation was also made in a separate study where the cells were expanded *in vitro* before being given back to the patient. The cells were harvested and expanded in the laboratory so that once reinfused, they would already be producing cells. In this study as well there were no “adverse events” associated with the transplant itself (Ikeda *et al.*, 2000). The relatively low numbers of transplant side effects are, however, greatly overshadowed by the direct toxic effects of the high dose chemotherapy.

Higher levels of chemotherapy induce higher levels of toxicity. Many studies indicated that a higher concentration of drugs and/or shorter treatment intervals were associated with higher levels of toxicity (Cossaart *et al.*, 2003; Genre *et al.*, 2002; Hortobagyi *et al.*, 2000; Stadtmauer *et al.* 2000). Genre *et al.* indicated that the need for hospitalization for this procedure was crucial, due to the potentially fatal side effects. However, it must be noted that Genre *et al.* were testing the limits of the human body’s tolerance. This study is therefore biased in that respect as evaluation of the safety limit necessarily created potentially unsafe treatment settings. In most studies reviewed, though, few actual toxic deaths were observed.

The harmful effects of the high dose chemotherapy itself, almost completely independent of the stem cell transplant, determine the treatment's overall safety and are studied intently. In the studies reviewed here, there were several common side effects associated with the HDC: nausea, myelosuppression (decrease in bone marrow activity), infection, mucositis, and diarrhea (Carlson *et al.*, 2001, Gaston-Johansson *et al.*, 1999, Wettergren *et al.*, 1999, Cossaart *et al.*, 2003, Hortobagyi *et al.*, 2000, Genre *et al.*, 2002). The patients' ages and receptor statuses were not isolated in these studies as limiting factors, making it difficult to deduce how these variables affect the safety of the treatment. Naturally, a younger woman will have a stronger immune system to help her recover from side effects.

The pulmonary circuit is also highly affected by chemotherapy, and in several studies, non-fatal heart related problems were noted (Hortobagyi *et al.*, 2000; Genre *et al.*, 2002; Cossaart *et al.*, 2003). *Herpes zoster* infections too were noted, along with fevers stemming from other infections (Greer *et al.*, 2002). Again, in most studies these infections were not fatal and the patient recovered fully. In all cases comparing levels of chemotherapy, the cohorts that received higher levels of drugs experienced more toxicity, but not necessarily toxic death.

According to one case report, ophthalmic disorders are another common complication. Cossaart *et al.* followed one woman through chemotherapy and stem cell transplant. Radiation therapy was also given due to the advanced stage of her metastatic breast cancer. Three weeks after the transplant, her vision became blurry and she experienced unsteadiness in walking. These problems progressed along with other neurological manifestations such as memory disturbance and hallucinations. Major ophthalmic disorders noted included retinopathy, optic neuritis and blindness, as well as blurred vision. It was concluded that high doses of the drugs given, specifically carmustine and carboplatin, damaged blood vessels in and around the brain, causing the blood-brain barrier to be breached (Cossaart *et al.*, 2003). No general conclusions can be drawn from this isolated case report, of course, and it is not generally known how common these ailments are. Nevertheless, any toxic side effects directly involving the central nervous system cannot be underestimated.

An important factor in judging the safety of HDC/ASCT is how the women feel during and after treatment and how they perceive their quality of life. Fatigue, pain, depression, and post-traumatic stress are as important as physical effects when considering the safety of this treatment. A study by

Gaston-Johansson *et al.* followed women who completed the chemotherapy and were awaiting a stem cell transplant. The patients were given questionnaires to fill out regarding their mental and physical symptoms. Duration, site and intensity of the ailments were recorded and correlations were drawn based on the answers received. This study is subjective, as every patient had a different level of pain tolerance and each judged herself in a unique manner (1999). Fatigue was the most commonly reported of all physical and mental symptoms of both the cancer itself and post-chemotherapy, according to ninety-one percent of the respondents. Perceptions that the women have of their own health can be skewed by fatigue (Gaston-Johansson *et al.*, 1999); furthermore, fatigue also can give way to other mental disturbances, such as depression, decreased sexual interest and mood instabilities.

The diagnosis of cancer is very intimidating and unfortunately, some women sink into a depression, which may be exacerbated by treatment, as the women worry about their futures and are forced to tolerate side effects. Carlson *et al.* showed that higher levels of depression were reported when women reported high amounts of pain, and levels were highest near the end of the patients' lives (Carlson *et al.*, 2001). Depression seemed to dissipate when overall health status improved, thus increasing the quality of life for some women (Gaston-Johansson *et al.*, 1999). Depression affects every aspect of one's life, which may mask any treatment benefits.

Stress is another factor affecting treatment outcomes because high levels of stress may directly impact patients; both their mental and physical health. A study on post-traumatic stress disorder in stem cell transplant patients confirmed harmful effects on patients' quality of life. This study was not limited to patients with metastatic breast cancer, but was completed in a similar way as the previously mentioned studies that recorded personal health. Patients were asked to fill out questionnaires to judge if they were suffering from post-traumatic stress as a result of stem cell transplant. It was reported that the week before the transplant, approximately half of the patients had the highest levels of distress, which declined only slightly after the transplant was completed (Wettergren *et al.*, 1999). The body's response to stress is complicated involving several organs, hormones and responses. This is extra energy that the body is expending and is therefore less able to focus on the task of recovery.

The overall quality of life in women with metastatic breast cancer varies over time for every patient and depends on the stage of treatment. It has been shown that all of

the factors that determine quality of life (hardship, pain, depression, physical symptoms, etc.) are more prevalent and more difficult to tolerate at certain times than others. Carlson *et al.* examined quality of life factors as the patients with metastatic breast cancer underwent high dose chemotherapy and then autologous stem cell transplant. The women filled out questionnaires measuring all aspects of life: levels of social functioning, physical pain, hardship, psychological stress, depression, anxiety, and appetite. The time when quality of life was lowest was when the women were near the end of their lives, as there are often high pain and feelings of hopelessness (Gaston-Johansson *et al.*, 1999). It was also observed that women who had a relapse of cancer showed lower quality of life when this was discovered. During the treatment, anxiety was the most frequently reported symptom, fueled by tiredness and depression (Carlson *et al.*, 2001). The use of HDC/ASCT affects women in many complex ways beyond cancer biology: it affects the whole person. These effects are generally not positive, but there have not been studies tracing the quality of life of patients not being treated for comparison.

Metastatic breast cancer is an often fatal and always life altering-disease with no ideal cure. The use of high dose chemotherapy with an autologous stem cell transplant is one of the options that doctors use today, but this combined treatment has many limitations when it comes to the safety of the 30,000 women who have chosen it. The physical and emotional trauma associated with chemotherapy is worsened by this more intense treatment. This research concludes that HDC/ASCT is generally safe for patients with metastatic breast cancer and is a viable treatment option when faced with no alternative therapy or is non-responsive to other treatments. While the mental anxieties and physical perils negatively affect the women who suffer from them, it is reasonable to assume that without this treatment, or indeed any treatment, these negative effects would be inflated.

Current developments have helped propel the safety and efficacy of HDC/ASCT. The practice of re-infusion of stem cells has become a very safe and effective method of hematopoietic rescue, as longer survival and few long-term illnesses are associated with it. Also, since this procedure is often done in an outpatient setting, it would seem that doctors have a high level of comfort with the procedure as it is currently done. There is a recommendation that the higher risk patients have the treatment in a hospital setting, as well as those using experimentally high intensity drugs. Helping women cope with cancer in addition to treating the

cancer would improve the quality of life and possibly the survival time that these women experience. Perhaps mental health counseling and other such support services would be useful adjuncts to any cancer treatment regimen.

Future Prospects

While there have been some studies that showed increases in progression free survival time and/or overall survival time in women with metastatic breast cancer who undergo HDC/ASCT, it appears as though the pool such candidates who may benefit is extremely limited. It may be that HDC/ASCT should be offered as only one component to a multifold treatment strategy in effort to remedy the cancer in the patient. More research is required to properly identify the specifics that will define the subgroup of women who will encounter these long-term advantages of HDC/ASCT.

While the outlook for the efficacy HDC/ASCT for metastatic breast cancer is not promising, there have been leads that this treatment may be more beneficial to patients with small tumors that are located in detectable sites (oligometastatic breast cancer). Results from a study by Nieto *et al.* suggest that a notable fraction of oligometastatic patients are likely to benefit from the HDC/ASCT treatment (2002). As most studies have solely focused on MBC and have encountered little success, in the future it would be advantageous to turn the research focus to oligometastatic patients when determining benefiting subgroup.

Furthermore, the development of drugs that would not harm healthy cells while killing cancerous cells would be the ideal solution to cancer. Researchers have been trying to find just such a solution for many years. Greater emphasis must continue to be placed on lessening the side effects of current treatment options. Support for the women who receive treatment should be augmented. Research into patient care and additional treatments for toxicity would also be beneficial, as well as psychological care for the patients.

Another issue that cannot be overlooked is money. This treatment can cost almost \$100,000, and not very many insurance companies are willing to cover the cost of the entire treatment since it is still technically in the experimental stages. Overall, contradictory studies have left HDC/ASCT as a treatment for metastatic breast cancer in limbo. The American Society of Clinical Oncology recommends that women only undergo this procedure if they are able to participate in a "high-quality clinical trial" (National Cancer Institute). Hopefully, the further advancement of

technology will make more treatment options with fewer side effects available to patients struggling with the deadly disease.

References

- Berry D, Broadwater G, Klein J, Antman K, Aisner J, Bitran J, Costanza M, Freytes CO, Stadtmauer E, Gale RP, Henderson IC, Lazarus HM, McCarthy PL Jr, Norton L, Parnes H, Pecora A, Perry MC, Rowlings P, Spitzer G, Horowitz MM. 2002. High-Dose Versus Standard Chemotherapy in Metastatic Breast Cancer: Comparison of Cancer and Leukemia Trial With Data From the Autologous Blood and Marrow Transplant Registry. *J. of Clinic. Oncol.* 20: 743-750.
- Byers T. 2000. High-Dose Chemotherapy for Breast Cancer? *Cancer Prac.* 8:96-98.
- Carlson LE, Koski T, and Gluck S. 2001. Longitudinal effects of high-dose chemotherapy and autologous stem cell transplantation on quality of life in the treatment of metastatic breast cancer. *B. Marrow Trans.* 27:989-998
- Cossaart N, Santacruz KS, Preston D, Johnson P, Skikne BS. 2003. Fatal chemotherapy-induced encephalopathy following high-dose therapy for metastatic breast cancer: a case report and review of the literature. *B. Marrow Trans.* 31:57-60.
- Dervan PA. 2001. *Understanding Breast Cancer.* McFarland & Company, Inc., Publishers, Jefferson, North Carolina – London.
- Folet JF, Vose JM, Armitage JO. 1999. *Current Therapy in Cancer.* WB Saunders Company, Philadelphia – London – Toronto – Montreal – Sydney – Tokyo.
- Gaston-Johansson F, Fall-Dickson JM, Bakos AB, Kennedy, MJ. 1999. Fatigue, Pain, and Depression in Pre-Autotransplant Breast Cancer Patients. *Cancer Prac.* 7:240-247.
- Genre D, Viens P, Bertucci F, Chabannon C, Gravis G, Braud A-C, Camerlo J, Houvenaeghel G, Moutardier V, Goncalvez A, Protiere C, Bardou V-J, Maraninchi D. 2002. Modulations of dose intensity of doxorubicin and cyclophosphamide in association with G-CSF and peripheral blood stem cells in adjuvant chemotherapy for breast cancer: comparative evaluation of completion and safety of three intensive regimens. *B. Marrow Trans.* 29:881-886.
- Greer M. 2002. CD34-Enriched Stem Cell Grafts Hasten Recovery After Myeloablative Chemotherapy. *Cancer Weekly.* Pg. 2.
- Hensel, M, Schneeweiss, A, Sinn, HP, Egerer G, Kornacker M, Solomayer E, Haas R, Bastert G, Ho AD. 2002. Stem Cell Dose and Turmobiologic Parameter as Prognostic Markers for Patients with Metastatic Breast Cancer Undergoing High-Dose Chemotherapy with Autologous Blood Stem Cell Support. *Stem Cells.* 20:32-40.
- Hortobagyi GN, Buzdar AU, Theriault RL, Valero V, Frye D, Booser DJ, Homes FA, Giralt S, Khouri I, Andersson B, Gajewski JL, Rondon G, Smith TL, Singletary SE, Ames FC, Sneige N, Strom EA, McNeese MD, Deisserogh AB, Champlin RE. 2000.
- Ikeda Y, Hata J, Koyasu S. 2000. *Cell Therapy.* Springer-Verlag, Tokyo, Japan: 120-122.
- Kim YS, Konoplev SN, Montemurro F, Hoy E, Smith TL, Rondón G, Champlin RE, Sabin AA, Ueno NT. 2001. HER-2/neu Overexpression as a Poor Prognostic Factor for Patients with Metastatic Breast Cancer undergoing High-dose Chemotherapy with Autologous Stem Cell Transplantation. *Clin. Cancer Res.* 7: 4008-4012.
- National Cancer Institute. High Dose Chemotherapy for Breast Cancer [Online]. <http://www.cancer.gov/clinicaltrials/developments/high-dose-chemo-history0501>. [4/30/2001].
- Nieto Y, Nawaz S, Jones R, Shpall E, Cagnoni P, McSweeney P, Barón A, Razook C, Matthes S, Bearman S. 2002. Prognostic Model for Relapse After High-Dose Chemotherapy With Autologous Stem-Cell Transplantation for Stage IV Oligometastatic Breast Cancer. *J. of Clinic. Oncol.* 20:707-718.
- Prucha EJ and Bellenir K (ed). 2001. *Breast Cancer Sourcebook.* Omnigraphics, Inc., Detroit, MI:56.
- Rowlings PA, Williams SF, Antman KH, Fields KK, Fay JW, Reed E, Pelz CJ, Klein JP, Sobocinski KA, Kennedy MJ, Freytes CO, McCarthy PL Jr, Herzig RH, Stadtmauer EA, Lazarus HM, Pecora AL, Bitran JD, Wolff SN, Gale RP, Armitage JO, Vaughan WP, Spitzer G, Horowitz MM. 1999. Factors Correlated With Progression-Free Survival After High-Dose Chemotherapy and Hematopoietic Stem Cell Transplantation for Metastatic Breast Cancer. *J. of the Amer. Med. Ass.* 282: 1335-1343.

- Schneeweiss, A, Hensel, M, Sinn, P, Khbeis, T, Hass R, Bastert, G, Ho, A. 2002. Characteristics associated with long-term progression-free survival following high-dose chemotherapy in metastatic breast cancer and influence of chemotherapy dose. *Annals of Oncol.* 13: 679-688.
- Schrama JG, Faneyte IF, Schornagel JH, Baars JW, Peterse JL, van de Vijver MJ, Dalesio O, van Tinteren H, Rutgers EJT, Richel DJ, Rodenhuis S. 2002. Randomized trial of high-dose chemotherapy and hematopoietic progenitor-cell support in operable breast cancer with extensive lymph node involvement: final analysis with 7 years of follow-up. *Euro. Soc. for Med. Oncol.* 13: 689-698.
- Schwartzberg LS, Weaver CH, Campos L, Tauer K, Smith R Jr., Zhen B, Birch R, Murphy MN, Buckner CD. 1999. High-Dose Chemotherapy with Peripheral Blood Stem Cell Support for Operable Locally Advanced Noninflammatory Carcinoma of the Breast. *Breast Journal.* 5:238-245.
- Stadtmauer EA, O'Neill A, Goldstein LJ, Crilley PA, Mangan KF, Ingle JN, Brodsky I, Martino S, Lazarus HM, Erban JK, Sickles C, Glick JH, Luger SM, Klumpp TR, Litzow MR, Topolsky DL. 2000. Conventional-Dose Chemotherapy Compared with High-Dose Chemotherapy plus Autologous Hematopoietic Stem-Cell Transplantation for Metastatic Breast Cancer. *New. Eng. Journal of Med.* 342:1069-1076.
- Thomas ED, Blume KG, Forman SJ (ed). 1999. *Hematopoietic Cell Transplantation.* Blackwell Science, Inc., Malden, MA:1033.
- Wettergren L, Langius A, Bjorkholm M, Bjorvell H. 1999. Post-traumatic Stress Symptoms in Patients Undergoing Autologous Stem Cell Transplantation. *Acta Oncologica.* 4:475-480.
- Williams, SF. 2002. Is There a Role for Dose-Intensive Chemotherapy With Stem Cell Rescue in Breast Cancer? *Oncology.* 16: 1643-1656.
- Winter JN. 1997. *Blood Stem Cell Transplantation.* Kluwer Academic Publishers, Boston-Dordrecht-London.

Natural Gas Sweetening Process Design

Kurt Mearkeltor

MSc in Chemical Engineering and candidate to PhD in Chemical Engineering at the
Department of Science of the School of Doctoral Studies of the EU

Email: k.mearkeltor@dsc.sds.eu

Abstract

For decades to come, gas will be the energy source of choice to meet worldwide environmental standards. Fortunately gas reserves are growing; but new gas is often found to be of substandard quality in remote and / or stranded areas of the world. When natural wellhead or oil field associated gases are highly loaded with acid gases, the dilemma facing most operators is what to do, how and when to best exploit these poor quality resources. Today the advanced activated MDEA process offers economy and versatility in handling both selective and complete acid gas removal services. The process has a good synergy with modern Claus sulfur recovery processes and remains among the best alternatives even when no sulfur recovery is foreseen. Nevertheless, there are limitations of even the most advanced amines only based gas treatment technologies in handling very highly acid gas loaded natural or associated oil field gases; especially for bulk acid gas removal when the acid gases are destined for cycling and / or disposal by reinjection. Today cycling and disposal by re-injection offers a promising alternative to avoid sulfur production and reduce CO₂ emissions to the atmosphere simultaneously. To this end, technologies of choice are those which offer maximum simplicity and require least downstream processing intensity and power for re-injection. **Keywords:** Chemical Engineering; Natural Gas Sweetening Process.

Natural gas from high-pressure wells is usually passed through field separators at the well to remove hydrocarbon condensate and water (Kirk & Othmer 1951). Natural gasoline, butane, and propane are usually present in the gas, and gas-processing plants are required for the recovery of these liquefiable constituents (Kirk & Othmer 1951).

Natural gas is considered “sour” if hydrogen sulfide (H₂S) is present in amounts greater than 5.7 milligrams per normal cubic meters (mg/Nm) (0.25 grains per 100 standard cubic feet [gr/100 scf]) (Maddox 1974). The H₂S must be removed (called “sweetening” the gas) before the gas can be utilized. If H₂S is present, the gas is usually sweetened by absorption of the H₂S in an amine solution (Maddox 1974). Amine processes are used for over 95 percent of all gas sweetening in the United States. Other methods, such as carbonate processes, solid bed absorbents, and physical absorption, are employed in the other sweetening plants. Emission data for sweetening processes other than amine types are very meager, but a material balance on sulfur will give accurate estimates for sulfur dioxide (SO₂) (Mullins 1975).

The major emission sources in the natural gas processing industry are compressor engines, acid gas wastes, fugitive emissions from leaking process equipment, and if present, glycol dehydrator vent streams. Fugitive leak emissions are detailed in *Protocol For Equipment Leak Emission Estimates* (EPA 1995). Regeneration of the glycol solutions

used for dehydrating natural gas can release significant quantities of benzene, toluene, ethyl benzene, and xylene, as well as a wide range of less toxic organics. These emissions can be estimated by a thermodynamic software model (*GRI-GLYCalc*), available from the Gas Research Institute.

Many chemical processes are available for sweetening natural gas. At present, the amine process (also known as the Girdler process) is the most widely used method for H₂S removal. The process is illustrated below:



Where:

- R = mono, di, or tri-ethanol
- N = nitrogen
- H = hydrogen
- S = sulfur

(Katz et al. 1959).

The recovered hydrogen sulfide gas stream may be: (1) vented, (2) flared in waste gas flares or modern smokeless flares, (3) incinerated, or (4) utilized for the production of elemental sulfur or sulfuric acid (EPA 1970).

For decades to come, gas will be the energy source of choice to meet worldwide environmental standards.

Fortunately gas reserves are growing; but new gas is often found to be of substandard quality in remote and / or stranded areas of the world. When natural wellhead or oil field associated gases are highly loaded with acid gases, the dilemma facing most operators is what to do, how and when to best exploit these poor quality resources.

Today the advanced activated MDEA process offers economy and versatility in handling both selective and complete acid gas removal services. The process has a good synergy with modern Claus sulfur recovery processes and remains among the best alternatives even when no sulfur recovery is foreseen (Blanc et al. 1981).

Nevertheless, there are limitations of even the most advanced amines only based gas treatment technologies in handling very highly acid gas loaded natural or associated oil field gases; especially for bulk acid gas removal when the acid gases are destined for cycling and / or disposal by reinjection (Elgue, Peytavy, & Tournier-Lasserre 1991). Today cycling and disposal by re-injection offers a promising alternative to avoid sulfur production and reduce CO₂ emissions to the atmosphere simultaneously. To this end, technologies of choice are those which offer maximum simplicity and require least downstream processing intensity and power for re-injection (Elgue, Peytavy, & Tournier-Lasserre 1991).

The behavior of CO₂ and H₂S in combination with water at high pressures and at temperatures within process operating conditions can result in hydrate formation, liquid phases of the compounds, and large variations in the amount of absorbed water. Since similar experimental data for mixtures of acid gases and water are not always readily available in the public literature, it is necessary to rely on computer generated results for mixtures for designing compression and injection facilities, which may not be entirely accurate (Royan & Wichert 1995).

Natural gas reservoirs have always been associated with water; thus, gas in the reservoir is water saturated. When the gas is produced water is produced as well. Some of this water is produced water from the reservoir directly (Sharma & Campbell 1969). Other water produced with the gas is water of condensation formed because of changes in temperature and pressure during production.

In the sweetening of natural gas, the removal of hydrogen sulfide and carbon dioxide, aqueous solvents are usually used (Sharma & Campbell 1969). The sweetened gas, with the H₂S and CO₂ removed, is saturated with water. In addition, the acid gas byproduct of the sweetening is also saturated with water. Furthermore, water is an interesting problem in the emerging technology for disposing of acid

gas by injecting into a suitable reservoir – acid gas injection (Sharma & Campbell 1969).

In the transmission of natural gas further condensation of water is problematic. It can increase pressure drop in the line and often leads to corrosion problems. Thus, water should be removed from the natural gas before it is sold to the pipeline company.

For these reasons, the water content of natural gas and acid gas is an important engineering consideration (Sharma & Campbell 1969).

Literature Review

Sour natural gas contains hydrogen sulfide (H₂S), which has to be removed to meet specifications for sales gas. Sour natural gas also contains carbon dioxide (CO₂). The removal of CO₂ and H₂S, usually called acid gases, from sour natural gas is generally accomplished by means of a regenerative solvent. There are several amine solvents used for this purpose. Upon regeneration of the solvent, the acid gases are liberated, and are usually sent to a modified Claus plant, where the H₂S is converted to elemental sulfur (Canjar & Manning 1967). The acid gas stream to the modified Claus plant consists of H₂S, CO₂, water vapor and minor amounts of hydrocarbon gas.

When the concentration of CO₂ is considerably greater than the concentration of H₂S in the acid gas mixture, the Claus plant has difficulty in achieving a high sulfur recovery. If the total sulfur rate is small, say less than 5 tons per day, it may be more economical to recover the sulfur by some other process. Such other processes, however, have many drawbacks of their own (Canjar & Manning 1967).

An alternative to recovering sulfur is to compress and reinject the acid gases into a suitable underground zone, in a manner similar to the disposal of produced water. An additional benefit of reinjecting the acid gases is the elimination of emission of sulfur compounds as well as CO₂ to the atmosphere (Canjar & Manning 1967).

Properties of H₂S and CO₂

Upon removal of the acid gases H₂S and CO₂ from the sour gas, an acid gas mixture is obtained at low pressure that may also contain about 1 % to 3 % hydrocarbon gases, and which is saturated with water vapor (West 1948). This is the mixture that is compressed through 4 stages of compression, from about 100 kPa (ga) to around 8 to 10 MPa. In this process, water condenses, creating the potential for corrosion and hydrate formation. In addition,

at such final compressor discharge pressures, the acid gas becomes a liquid or a dense phase when cooled to ambient temperatures (West 1948).

While experimental results of studies of the physical properties of acid gas mixtures without hydrocarbon components are difficult to determine in the technical literature, the properties of pure CO₂ and H₂S have been examined and reported (West 1948).

Additionally, the properties of each of the acid gases have also been studied in the presence of water at elevated pressures and temperatures.

These results can be used as a guide to indicate how the mixed acid gas streams would behave under the conditions of pressure and temperature when compressed to the injection pressure level.

A brief review of the properties of the pure acid gases and the CO₂ - water and H₂S - water binaries is therefore appropriate.

Vapor / Liquid Properties of Pure Compounds

In their pure state, CO₂ and H₂S exhibit the normal vapor / liquid behavior with pressure and temperature. At higher pressure and temperatures, the pure compounds exist as liquid and vapor. Methane (CH₄) also exhibits this behavior, but at much lower temperatures. From this it can be seen that CO₂ and H₂S can readily be liquefied at elevated pressures and at relatively high temperatures, whereas CH₄ can be liquefied only at very low temperatures (Song & Kobayashi 1989).

Mixtures of the acid gases are also readily liquefied at elevated pressures, but not mixtures containing substantial amounts of CH₄ in their makeup.

CO₂-Water and H₂S-Water Mixtures

When a mixture of CO₂ or H₂S and water is subjected to elevated pressures, another phase, namely a solid hydrate phase, is formed under certain conditions of temperature and elevated pressures. Hydrates can form at elevated pressures and at fairly high temperatures with H₂S, and at somewhat lower temperatures with CO₂ (Song & Kobayashi 1989). These pressures and temperatures are usually encountered in acid gas injection processes.

Water of Saturation for CO₂ and H₂S

The ability of the pure compounds to hold water in the vapor phase is reduced as the pressure increases at about

2.7 MPa for H₂S and up to about 6 MPa for CO₂. When pressures are raised above these levels, the capacity of H₂S and CO₂ to hold water in solution increases substantially, as both compounds have a higher water absorption capacity in the liquid phase or dense phase as compared with their vapor phases (Song & Kobayashi 1989). A small amount of methane reduces substantially the water absorption ability of CO₂ at elevated pressure.

Other water content data at various temperatures and pressures for CO₂ and H₂S are illustrated in Figures 20-4 to 20-7 of the *GPSA Engineering Data Book*, on page 20-5 of the 1994 edition.

Acid Gas Mixtures

It is well known in the natural gas industry that raw sour natural gas mixtures present increased problems as compared to sweet natural gas in mainly two areas: increased potential for corrosion and increased propensity for the formation of hydrates at elevated pressures. It is therefore assumed that acid gases with little or no hydrocarbon content will further exacerbate these two problems. In actual fact, this may not be the case.

Problem of Corrosion

Corrosion of low alloy steel by sour gas occurs mainly in the presence of a liquid aqueous phase. The corrosion manifests itself as general metal loss corrosion or pitting corrosion (Selleck et al. 1952). When these occur, atomic hydrogen is generated, which has the ability to penetrate steel. This can lead to hydrogen induced cracking, blistering or sulfide stress-cracking, resulting in sudden failures of the confining pipes or vessels. By following the *NACE MR0175* standard in materials selection and adhering to recommended construction practices, such failures can usually be avoided.

By dehydrating sour gas, corrosion by H₂S or CO₂ is effectively eliminated. Similarly, the potential for corrosion by acid gas can also be controlled by dehydration.

Problem of Hydrates

The hydrate forming temperature at any pressure of a sour natural gas increases with increasing content of H₂S. The opposite is the case with CO₂. For sour natural gases with low concentrations of H₂S, the concentration of CO₂ is usually much greater than the concentration of H₂S (Wiebe & Gaddy 1941). Thus the acid gas mixture, upon

extraction from the sour gas, would be mainly CO_2 . Hydrate temperatures at H_2S contents below 5 % are no higher than what would be the case for sweet natural gas. Hydrates of course will not occur if the gas has no liquid water phase.

Vapor / Liquid Phase Behavior

Any gas can be liquefied if it is sufficiently cooled. This means that aerial cooling readily liquefies acid gas mixtures at elevated pressures. If the acid gas mixture contains considerable amounts of light hydrocarbon gas, then the two-phase envelope for the mixture becomes wider in temperature – an example being a mixture that contains 10 % CH_4 and 45 % each of H_2S and CO_2 (Song & Kobayashi 1984).

To completely liquefy such mixtures requires somewhat lower temperatures. If the acid gas mixture is ultimately to be liquefied upon compression and aerial cooling, then the sour gas treating plant must be designed and operated to minimize the hydrocarbon gas content of the acid gas mixture. The design should include a rich solution flash tank, and this vessel should be operated at a low pressure so that most of the dissolved hydrocarbon gas can be liberated. The solvent should be operated at the upper range of solvent concentration, and solution circulation should be such that the mole loadings (moles of acid gas picked up in the contactor per mole of solvent circulated) are also at the upper range of the recommended loadings (Song & Kobayashi 1989).

Water Vapor Content

A high hydrocarbon content also adversely influences the water carrying capacity of the acid gas mixtures at high pressures. The equilibrium water vapor content of natural gas decreases with increasing pressure at constant temperature. Similarly, in their pure state, both H_2S and CO_2 hold less water in the vapor phase as the pressure is increased. However, in the liquid or dense phases of H_2S and CO_2 , the water of saturation increases substantially with increasing pressure at a given temperature, due to the molecular attraction between these polar compounds (Song & Kobayashi 1984).

It is reasonable to assume that the behavior of mixtures of H_2S and CO_2 would be similar to their behavior in the pure states. If this assumption is correct, then the acid gas mixtures without hydrocarbon content will also show a minimum water content at a pressure of about 3000 to 5000

kPa, depending on the relative amounts of H_2S and CO_2 in the mixtures. Thus, by separating the condensed water in this pressure range and at a temperature a few degrees above the hydrate temperature, no free water should be present upon compressing the acid gases to substantially higher pressures and cooling to ambient temperatures (Song & Kobayashi 1984). At such elevated pressures, the acid gas mixture would be in the liquid phase or the dense phase at normal pipeline temperatures. This means that there may not be any need to dehydrate the acid gas mixture, as no liquid water phase would form after the final stage of compression and upon cooling to ambient temperatures. Without liquid water present, no hydrates would occur in the liquid phase of the acid gas mixture, and corrosion would be at a minimum (Selleck et al. 1952).

Acid Gas Compression and Dehydration

The acid gas is liberated from the sweetening solvent in the regenerator tower, and this overhead vapor stream is then cooled. Most of the water and solvent is condensed in the cooler, and the two-phase mixture flows to the reflux drum, where the condensed liquids are separated and returned to the regenerator column as reflux. The vapor stream leaving the reflux drum contains the acid gases, H_2S and CO_2 , a small percentage of hydrocarbon gas and some water vapor (Wiebe & Gaddy 1941).

The pressure of this gaseous mixture is generally in the order of 80 to 100 kPa (ga), and at a temperature between 20°C in winter and up to 40°C in summer. This is the stream that would normally be sent to a flare stack if the sulfur content is less than 1 t/d, or be sent to a sulfur recovery unit if the sulfur content is 1 t/d or greater. If the acid gas contains more than say 60% CO_2 , which is usually the case with small amounts of H_2S in the sour natural gas, then a modified Claus plant may not be the best choice for small-scale sulfur recovery, and instead a redox unit could be selected to convert the H_2S to sulfur (Song & Kobayashi 1989). However, redox processes are expensive to install, difficult to operate, and the resulting sulfur product does not meet the sulfur sales specifications.

Compressor Discharge Pressure

The ultimate pressure to which the acid gas has to be compressed depends on the pressure of the reservoir, the permeability of the zone, and the depth of the zone. If the acid gas mixture is compressed to about 6000 kPa and

cooled below 20°C, the mixture will be in the liquid phase, provided the methane content is no greater than one or two percent (Wiebe & Gaddy 1941). The density of a liquid acid gas mixture at such conditions will have a density of about 70 to 80 % of the density of water. Thus the density of the acid gas stream in the liquid state will aid the injection pressure into the reservoir. Also, the rate of injection will be generally low, less than 1 Bbl/min. One Bbl/min. of liquid CO₂ injection amounts to about 100 10₃m₃/d or 3.6 MMscf/d of gaseous CO₂ at standard pressure and temperature. Many zones have a pressure near the hydrostatic head of water (Selleck et al. 1952).

Thus the wellhead injection pressure would generally be in the order of 6000 to 9000 kPa, and would depend largely on the injectivity of the reservoir as well as reservoir pressure and depth. A four – stage compressor, can achieve the pressure of 6000 to 9000 kPa.

Table 1 shows the typical operating pressures for a four – stage compressor at a compression ratio of 3 between stages, and Figure 5 presents the variation in temperature between stages for an acid gas mixture containing 1% CH₄, 49% C₂ and 50% H₂S. The hydrate temperature with increasing pressure is shown with dashes.

Table 1. Typical Interstage Pressures for a Four-Stage Compressor

Stage No.	Suction P, kPa (abs.)	Discharge P, kPa (abs.)
1	150	455
2	385	1160
3	1090	3270
4	3200	9600

(Song & Kobayashi 1984)

The total horsepower requirement in the above example is about 7.94 kW per 10₃m₃/d (300 HP per Mmscf/d).

Cooler After Final Compression Stage

If the acid gas is to be cooled after the final stage of compression so that it is converted into a liquid or dense phase, the final cooler section has to be designed to remove the equivalent of the latent heat of condensation of the acid gas. In Figure 5, for example, the heat removed during cooling after the fourth stage is about four times that removed after the third stage.

Metallurgy

The compressor cylinders will compress gas that is at its water dew point on the suction side, and under-saturated in the compressor cylinder and on the discharge side, due to the temperature rise as a result of compression. Thus carbon steel meeting *NACE MR0175* standard requirements for sour gas should be installed as a minimum. The line to the aerial cooler can be carbon steel meeting the *NACE* standard. In the inter-stage coolers, water will condense, and could pose a corrosion problem. Downstream of the coolers, the lines to the inter-stage scrubbers and the scrubbers themselves will be exposed to the corrosive acid gas and condensed water mixture. Carbon steel meeting *NACE* specifications for sour gas, coupled with an appropriate corrosion inhibition program, should be all that is required to handle these corrosive fluids.

After the third stage of compression, the pressure will be in the order of about 3000 to 4000 kPa. This is the final stage at which water will condense in the cooler. Upon boosting the pressure substantially above 6000 kPa, the acid gas mixture can hold more water in solution than at the previous separation stage. Thus upon cooling after the fourth compression stage, there should theoretically be no water of condensation dropping out (Song & Kobayashi 1989). If this is so, then there would be no need to dehydrate the gas, and the downstream facilities could be constructed out of carbon steel meeting *NACE* specifications, which would save the gas industry a lot of money. This aspect of acid gas phase behavior should be experimentally investigated.

If dehydration is chosen to ensure that no water of condensation drops out at the high pressures, the dehydration should take place after the second or third stage of compression. When dehydration occurs after the second stage, the pressure is relatively low, but the acid gas mixture contains more water than after the third stage compression and cooling. With dehydration occurring after the third stage, the operating pressure of the dehydrator would be about 3 times the pressure after the second stage of compression. This would require less glycol circulation, smaller vessel diameter, but the amount of acid gases absorbed per gallon of glycol circulated would be higher. However, glycol circulation would need to be about one half of the required circulation for dehydration after the second stage of compression, since there is less water to remove from the acid gas stream at the higher pressure (Selleck et al. 1952).

In general, the selected metallurgy would depend on whether the gas is dehydrated or not. Stainless steel equipment includes the suction scrubbers, pulsation bottles, interconnecting piping and cooler tubes. The metallurgy of valves and instrumentation such as block valves, dump valves, PSV's, gauge glasses, etc., should generally match the material of the vessel and piping.

Compressor cylinder material can be carbon steel. Compressor suction and discharge valves should be stainless steel. The piston rod should be stainless steel, and may be coated with a tungsten carbide overlay having a maximum hardness of RC22. Distance pieces can be single compartment with a vacuum pump, or a two-piece compartment with a sweet gas purge system (Royan & Wichert 1995).

Because of the low suction pressure in the 1st stage suction scrubber, any liquids must be dumped to an atmospheric tank, or handled by a pump. Any liquids condensing in the remaining suction scrubbers are usually dumped to a sour water treatment and disposal system. Ideally, these liquids should be returned to the amine system, to save on make – up water. However, due to potential contamination with compressor oil, it is safest to discard the water that condenses in the suction scrubbers (Royan & Wichert 1995).

Compressor Units can be equipped with separate cylinder lubrication and flushing oil systems so that lube oils and corrosion inhibitors can be injected separately. Corrosion allowance in piping generally should be 3 mm in carbon steel materials and 1.5 mm in stainless steel materials (Royan & Wichert 1995).

Acid Gas Dehydration

Whether or not acid gas dehydration is necessary is debatable. As mentioned above, under certain conditions of pressure and temperature, no water of condensation would occur after the final stage of compression (Canjar & Manning 1967).

If dehydration is installed, then it should be part of the process at pressures of minimum water content in the acid gas stream. At a temperature of about 35 °C after cooling, the optimum pressure range is 3000 to 5000 kPa, depending on the acid gas composition. This is the pressure range in which acid gas mixtures should exhibit a minimum equilibrium water content at the process operating temperature range (Song & Kobayashi 1984).

The dehydration equipment would be a normal glycol unit, with a minimum of six trays, using triethylene glycol.

In sales gas dehydration, the water content has to be reduced to about 65 mg/m³ (4 lbs/MM). Since the objective in acid gas dehydration is simply the avoidance of a liquid water phase, the amount of water left in the acid gas can be substantially higher than 65 mg/m³. A safe water content would be about 500 to 700 mg/m³ (0.5 to 0.7 kg/103m³) (Song & Kobayashi 1984). Thus, the dehydration facilities for acid gas could be operated more economically than for sales gas dehydration.

One of the problems with dehydration of gas in general and with acid gas in particular is the disposal of the reboiler off-gases (Wiebe & Gaddy 1941). There is considerable environmental concern with venting these off - gases into the air. In acid gas dehydration, these off-gases have to be collected and incinerated or cooled and recompressed into the acid gas steam.

Acid Gas Injection Facilities

The acid gas injection facilities beyond the final stage of compression and cooling consist of an injection line, well-site control facilities and the injection well.

The Disposal Line

Selection of the type of pipeline material for the acid gas injection line between the plant and the injection well is generally related to whether or not the acid gas has been dehydrated. For dehydrated gas, sour service carbon steel materials can be used, such as CSA-Z662 Grade 359 Cat. II sour service material. The addition of corrosion inhibitor should be considered.

For non-dehydrated acid gas, the line can be of carbon steel with an internal epoxy coating, or a polyethylene liner (EPA 1995).

Another pipeline material can be 316L stainless steel. One of the key considerations is the length of the line. Stainless steel is more expensive, and if the cost of the line exceeds the cost of a dehydration unit, then it may be more economical-to dehydrate the acid gas stream. The proposed pipeline should be fully evaluated with the vendor of the materials to ensure compatibility with anticipated process conditions of pressure, temperature and acid gas composition (EPA 1995).

The line diameter should be sized for liquid phase fluids if the acid gas mixture contains no more than about 1 % hydrocarbons (EPA 1995). This means that the injection rates would be quite low in terms of liquid quantities.

The following formulae relate acid gas injection rates at standard conditions to liquid disposal volumes per minute:

$$q_1, \text{ liters/min} = Q, \text{ acid gas } 10^3 \text{ m}_3/\text{d} * (1.57 * \text{mol fr CO}_2 + 1.27 * \text{mol fr H}_2\text{S})$$

$$q_1, \text{ US gal/min} = Q, \text{ acid gas MMscf/d} * (11.8 * \text{mol fr CO}_2 + 9.34 * \text{mol fr H}_2\text{S})$$

(EPA 1995).

The superficial liquid velocity in the pipeline should be kept below 3 m/sec (10 ft/sec). Liquid velocity can be estimated by:

$$v, \text{ m/sec} = 21.2 * \text{liter/min} / (d_2), \text{ where } d \text{ is the internal diameter of the pipe, in mm}$$

$$v, \text{ ft/sec} = 0.409 * \text{US gal/min} / (d_2), \text{ where } d \text{ is the internal diameter of the pipe, in inches}$$

(EPA 1995).

From the above it can be seen that an injection rate of 28 $10^3 \text{ m}_3/\text{d}$ (1 MMscf/d) of CO_2 reduces to 44 L/min (11.8 US gal/min) in the liquid phase. Thus the pressure drop in a short, 50 mm diameter line would be quite small, at a liquid velocity of about 0.37 m/sec. If the disposal well is near the plant site, a 50 mm disposal line out of 316L SS would cost about \$ 40,000 (EPA 1995).

Under normal operating conditions the line between the plant and the injection well would contain the acid gas mixture in the liquid phase or the dense phase. The temperature of the fluid in the line would be at ground temperature, which could be as low as 0 °C in the winter. If this line had to be depressurized, however, very low temperatures would result in the line. Since the fluid would be blown down to atmospheric pressure, the temperature of the fluid in the line would decrease to the boiling point temperature of the acid gas mixture at atmospheric pressure (EPA 1995). This could be as low as - 50 °C, or lower if the line were insulated. Lines that are not insulated would benefit from heat transfer from the surrounding soil. The pipeline material has to be designed for low temperatures, even though under normal conditions such low temperature would not be reached. A safe blow-down design and procedure should be incorporated in such a scheme (EPA 1995).

The extremely low temperatures can be avoided by installing a 25 mm carbon steel line to the well site in the trench with the 50 mm injection line. The design

pressure of this sweet gas line would be 7 MPa, and under normal circumstances, this line would be left idle at a low pressure. The line would be connected at the well site to the injection line upstream of the wellhead (EPA 1995). If it became necessary to depressurize the injection line, the acid gas would be flowed back to the plant and into the flare system through a standby heat exchanger or a rental line heater. The liquid acid gas mixture would be heated and vaporized prior to the pressure reduction into the plant flare system. To prevent the acid gases from vaporizing in the injection line, sweet gas from the plant would be flowed to the well site through the 25 mm line, from where the sweet gas would push the acid gas fluids back to the plant at full pressure (EPA 1995).

Well site Facilities

At the well site, the facilities would be very simple. A meter should be installed to record the flow. This would provide information to the plant SCADA system concerning the injection pressure, temperature, rate and the fluid density. The wellhead would be equipped with a check valve as well as an ESD valve (EPA 1995).

The Injection Well

When an acid gas disposal scheme by downhole injection is considered, one of the first questions that arises is whether there is a suitable injection well within a reasonable distance, and the second question is whether the AEUB will permit injection of the acid gas into the particular zone that the well is completed in. In general, the Board will allow injection into the zone from which production is being taken provided the operator gives reasonable assurance that the injected acid gas will not adversely affect offsetting wells or hydrocarbon recovery (EPA 1995). A well some distance from the plant may not be the best choice, however. Firstly, it usually is not a part of the production scheme to the plant because it missed the producing horizon, or it is so tight that it is incapable of production. In such cases it may not be a suitable candidate for acid gas injection either.

The ideal acid gas injection well would be a well drilled for this purpose within 200 m of the plant perimeter. The zone chosen for acid gas injection should be a zone that contains salt water, which means that the zone in the plant area has no commercial value. Such zones tend to have good thickness, a high permeability and a good aerial extent (EPA 1995).

The drilling and completion costs are largely dependent on the depth of the horizon. Table 2 shows the cost estimates for drilling acid gas injection wells to 1500 and 2500 m in Alberta, and completing them with 73 mm J55 EUE tubing that is internally coated for corrosion protection.

Table 2. Cost Estimate To Drill And Complete Acid Gas Disposal Well In Alberta

Depth, m	Drill & Case	Complete	Total Cost
1,500	\$350,000	\$150,000	\$500,000
2,500	\$550,000	\$250,000	\$800,000

(EPA 1995).

The completion would include a downhole check valve, for safety reasons in case of surface equipment failure. A downhole packer would protect the casing and a non-corrosive fluid, such as diesel fuel containing corrosion inhibitor, in the annulus between the casing and tubing, as is normally required for water injection or disposal wells.

Cost Comparisons with Small Scale Sulfur Recovery Options

In a companion paper, the various options for small-scale sulfur recovery are discussed (Royan & Wichert 1995). The capital cost of any of sulfur recovery process depends, of course, on many variables, such as the plant design throughput, composition, location, etc. The comparative costs can best be illustrated by an example:

Estimate the capital cost of the sweetening facilities and the various sulfur recovery and disposal facilities for a sour gas plant designed to treat 500 10^3 m³/d of gas with 2.0 % CO₂ and 0.44 % H₂S.

(Royan & Wichert 1995)

The total sulfur content in the inlet gas stream in the above example is 2.98 t/d. The acid gas stream will contain 18 % H₂S, unless some CO₂ is left in the sweet gas by the chosen solvent. For the purpose of this example it is assumed that all of the CO₂ will be extracted. The capital cost estimate for the equipment, without installation in the field, is shown in Table 3.

Table 3. Estimate of Capital Cost of Various Sulphur Recovery Processes (Thousands of Dollars)

Process Equipment	Two-stage Claus	One-stage Selectox	LO-CAT II	SulFerox	Acid Gas Injection
Sweetening	650	650	650	650	650
Sulfur recovery:					
Facilities	1,000	1,000	2,400	1,800	0
Catalyst	6	40	0	0	0
S Filter/Melt	0	0	300	300	0
Solvent	75	75	300	300	0
Incin./Stack	150	150	0	0	0
Acid gas Injection:					
Acid gas compr.	0	0	0	0	650
Acid gas line (316SS)	0	0	0	0	40
Injection well	0	0	0	0	700
Royalties	0	100	150	200	0
Estimated Total	1,881	2,015	3,800	2,850	2,040

Note: Estimated capital costs are for skid mounted equipment f.o.b. Calgary shop.

(Royan & Wichert 1995)

The capital cost changes somewhat with the total plant throughput and the H₂S content. If in the above example the total sour gas inlet rate to the plant were 900 10^3 m³/d, having a CO₂ content of 4 % and a H₂S content of 0.12%, the capital cost estimate would be as per Table 4. In this case the sulfur inlet rate to the plant is 1.46 t/d, with the H₂S

concentration in the acid gas stream being 2.91 % (Royan & Wichert 1995).

To employ the modified Claus process in the second example requires that a process be added between the sweetening step and the Claus plant to enrich the acid gas sufficiently in H₂S so that a modified Claus unit will operate satisfactorily.

Table 4. Estimate of Capital Cost Of Various Sulphur Recovery Processes (Thousands Of Dollars)

Process Equipment	Two-stage Claus	One-stage Selectox	LO-CAT II	SulFerox	Acid Gas Injection
Sweetening	1,250	1,250	1,250	350	1,250
Sulfur recovery:				(See note 2)	
Acid gas enr.	750	0	0	0	0
Facilities	750	800	1,900	1,550	0
Catalyst	5	20	0	0	0
S Filter/Melt	0	0	300	300	0
Int. Chem. Chg.	225	75	250	250	0
Incin. /Stack	150	150	0	0	0
Acid gas Injection:					
Acid gas compr.	0	0	0	0	1,000
Acid gas line (316SS)	0	0	0	0	40
Injection well	0	0	0	0	700
Royalties	200	100	150	200	0
Estimated Total	3,300	2,395	3,850	2,650	2,990

Note 1. Estimated capital costs are for skid mounted equipment f.o.b. Calgary shop

Note 2. 4% CO₂ left in sales gas

(Royan & Wichert 1995)

Each of the sulfur recovery processes would release the CO₂ to the atmosphere, amounting to 6,800 and 24,500 tons annually for the two cases, respectively.

Operating Costs

Operating cost estimates for all processes are given in Table 5 and include of the following items:

- Utilities (electricity, fuel gas, water)
- Chemicals (catalyst, chelates, iron oxide ions, pH control, corrosion inhibitor)
- Maintenance
- Operator time

Table 5. Estimate Of Annual Operating Cost Of Various Sulphur Recovery Processes (Thousands Of Dollars, For Example 1 Above)

Cost Item	Two-stageClaus	One-stage Selectox	LO-CAT II	SulFerox	Acid Gas Injection
Electr. Power	76.3	95.5	78.4	104.0	153.3
Fuel gas	7.50	75.0	75.0	0	75.0
Catalyst	2.0	13.0	0	0	0
Chemicals	25.0	25.0	320.0	350.0	25.0
Labor, Op/Mnt	150.0	150.0	250.0	250.0	200.0
Sulfur Disp.	0	0	50.0	50.0	0
Estimated Total	328.3	358.5	773.4	754.0	453.3
\$ Per ton of S	301	330	711	693	416

Assumptions going into these cost estimates:

- Catalyst change-out: once per 3 years
- Electricity: 10 c/kWh
- Fuel gas: \$1.50/GJ
- Chemicals: routine consumption

(Royan & Wichert 1995)

The above cost estimates are for the first example, and include operator time. The costs for sulfur recovery schemes will not vary that much between the two examples, as the overall amount of sulfur recovered is not that large to begin

with. Regarding acid gas compression, however, the costs would be higher in the second example by about \$155 thousand in light of more compression for the higher acid gas rate.

Optimum Acid Gas Injection Scheme

In light of the above discussions, it is possible to project what a minimum acid gas compression and injection scheme would include – as discussed below:

Injection Well

The injection well should be drilled specifically for the purpose of acid gas injection. It should be located near the plant, so that the injection line would be short, and it would be within the Emergency Planning Zone of the plant. This would minimize the cost of the pipeline. The selected zone should be an aquifer of large extent, with good permeability. This means that the gas would not be reproduced, as would potentially be the case if the gas were injected into a hydrocarbon-bearing zone (EPA 1995).

The depth of the well should be between 1000 and 1500 m, so that the injection pressure did not have to be above 8000 kPa. This pressure level would ensure that the acid gas mixture is in the liquid phase upon final cooling, which would aid with the injection into the zone.

The well could be completed with 114.3 mm casing string and a 60.3 mm OD J-55 tubing string, as the liquid injection rate would in all likelihood be less than 100 L/min (27 US gal/min), thereby saving on drilling and completion costs. A subsurface check valve should be installed in the tubing (EPA 1995).

The Injection Line

Since the well would be located near the plant, the injection line would be relatively short. A 316L stainless steel line, 50 mm in diameter, would likely be the best choice. This line would need to be externally coated for corrosion protection, but would not need to be insulated (EPA 1995).

Under normal operating conditions, the acid gas liquid or dense stream would not cool sufficiently to have water drop out to cause problems. If the line were shut down for some time, the contents of the line would cool down to ground temperature at burial depth, which would generally not be below 0 °C (EPA 1995).

Any water that might drop out would be minimal, would form a hydrate, and would immediately be reabsorbed upon startup of injection. A small, high pressure, sweet gas line should be included from the plant to the well, for reasons explained above.

Water Removal

It is evident that free water cannot be tolerated at the injection pressure levels of the acid gas mixture, because hydrates could form. There is a minimum water of saturation content at certain pressure and temperature conditions. By taking advantage of this situation, it may be possible to lower the water content sufficiently by cooling the acid gas stream, so that dehydration would not be required (EPA 1995).

A more economical method than dehydration would be to cool the gas stream to about 30 or 35 °C after the third stage of compression. The cooling could be done with an aerial cooler to about 3 or 4 degrees above the hydrate temperature for the particular acid gas composition. This would condense most of the water while the acid gas mixture is in the vapor phase (EPA 1995). Upon compression and cooling after the fourth stage of compression, the acid gas stream would be in the liquid or dense phase. The water of saturation at the temperatures in the short line to the injection well would always be higher than the water content at the final separation stage. It is assumed in this example that the hydrocarbon content of the acid gas mixture is less than about 3 %. Higher concentrations of C_{1+} will adversely affect this acid gas-water relationship (EPA 1995).

Acid Gas Compression

An estimate has to be made of the final pressure to which the acid gas stream has to be compressed. This depends mainly on the formation pressure, which depends on the well depth in undisturbed reservoirs. Other factors influencing the injection pressure are the permeability of the zone, the thickness of the zone and the rate of injection. Assuming that the necessary pressure is between 6 MPa and 10 MPa, a four-stage compressor should be used (EPA 1995).

Concerning the metallurgy of the compressor, the suction scrubbers, the suction and discharge lines and the coolers, some items may need to be stainless steel, but with research in the area of water content, some of the lines could probably be constructed with carbon steel meeting *NACE* specifications for sour gas applications.

Conclusions

- The compression and reinjection of acid gas into an underground reservoir is technically feasible and economically viable, when compared with alternate methods of sulfur disposal.
- Research should be carried out to determine the water of saturation of acid gas mixtures containing up to 1% to 3 % methane, in the pressure range of 3 MPa to 10 MPa, and at temperatures between 0 °C and 40 °C. The results would establish whether the acid gas would need to be dehydrated to avoid the formation of hydrates under operating conditions of acid gas injection systems.
- Reinjection acid gas into an underground formation releases the minimum of greenhouse gases into the atmosphere when compared with the options for small-scale sulfur recovery.
- Of the various options for handling small volumes of acid gas, compression and injection into an underground zone provides the greatest flexibility in operation with changing throughput rate or fluctuating acid gas composition.

The Sharma – Campbell Method for Predicting Water Content

Sharma and Campbell propose a method for calculating the water content of natural gas, including sour gas. Although originally designed for hand calculations, this method is rather complicated. It is even rather complicated for computer applications.

The method will be described here. Given the temperature and the pressure, the procedure is as follows. Determine the fugacity of water at the saturation conditions (T and P_{sat} water), which is designated f_{sat water} and the fugacity at the water conditions (T and P_{total}), designated f_{water}. A chart is provided to estimate the fugacity of water at the system conditions. Then the correlation factor, k, is calculated from the following equation:

$$k = \frac{P_{sat} \text{ water} \cdot f_{sat} \text{ water}}{P_{total} \cdot f_{water}} \cdot \frac{P_{sat} \text{ water}}{P_{total}} \cdot P_{total}^{0.0049}$$

(Sharma & Campbell 1969)

In this equation a consistent set of units should be used for the pressure and fugacity terms and then k is dimensionless. Then you must obtain the compressibility factor (z-factor), z, for the gas again at system conditions.

Sharma and Campbell recommend using a generalized correlation for the compressibility. Finally, the water content is calculated as:

$$w = 47484 k \frac{f_{sat} \text{ water}^5}{f_{gas}}$$

(Sharma & Campbell 1969)

where f is the fugacity of the dry gas calculated at system conditions. Again, if a consistent set of units is used for the fugacity terms, then the calculated water content, w, is in lb/MMCF.

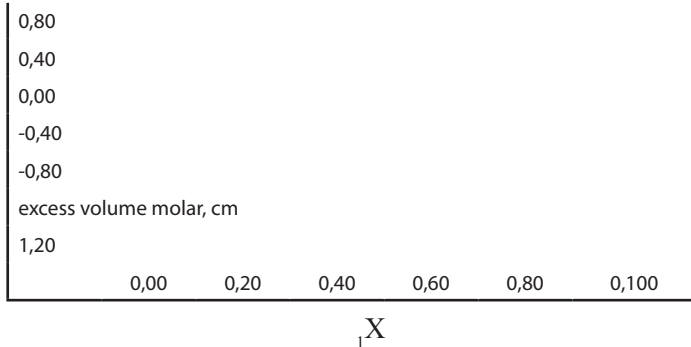
This method is rather difficult for hand calculations. First, it requires the compressibility factor of the gas mixture. Next, it requires the fugacity of pure water at system conditions. The chart given to estimate this value is only valid for temperatures between 80 and 160 degrees F, and for pressure less than 2000 psia (Sharma & Campbell 1969). It is unclear how this method will behave if extrapolated beyond this range. Typically, the calculation of a single fugacity is enough to scare away most process engineers. The Sharma – Campbell method requires three fugacity calculations for a single water content estimate.

The temperature and pressure limitations make this method less useful for some studies, where the pressure of interest usually ranges up to 10,000 psia and temperatures to 220 degrees F (Sharma & Campbell 1969). In addition, although this is intended to be a hand calculation method, it is a bit difficult to use.

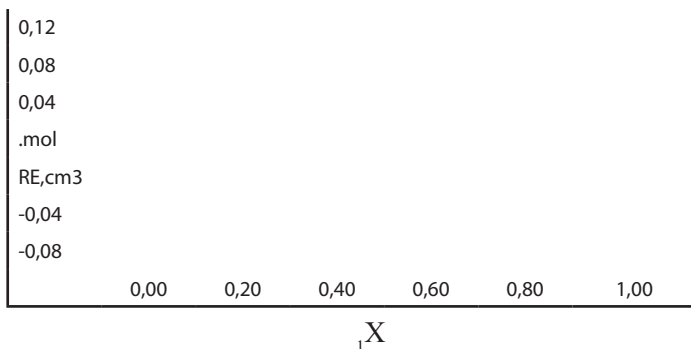
The Eykman Molecular Refraction (EMR) Mixture Combination Rules

For an example of the prediction of refractive indices, Eykman’s mixing rules are displayed below:

	A0	A1	A2	A3	A4	s
2-butanone (1)+1-chloropentane (2)						
$V_E, \text{cm}_3.\text{mol}_-1$	2,64654	0,13115	-0,00177	0,06722	0,00270	
$R_E, \text{cm}_3.\text{mol}_-1$	0,36011	-0,03642	-0,00717	0,00785	0,00100	
2-butanone (1)+1,3-dichloropropane (2)						
$V_E, \text{cm}_3.\text{mol}_-1$	-0,49936	0,98856	1,31123	0,25639	0,00234	
$R_E, \text{cm}_3.\text{mol}_-1$	0,16256	0,19454	0,11474	0,09188 -0,15178	0,00069	
2-butanone (1)+1,4-dichlorobutane (2)						
$V_E, \text{cm}_3.\text{mol}_-1$	-1,11593	0,72989	0,05418		0,00280	
$R_E, \text{cm}_3.\text{mol}_-1$	0,28129	0,42182	-0,46736	0,02022 -0,10627	0,00074	
2-butanone (1)+1,1,1-trichloroethane (2)						
$V_E, \text{cm}_3.\text{mol}_-1$	0,35307	0,62942	0,15189	0,04568	0,00009	
$R_E, \text{cm}_3.\text{mol}_-1$	0,12084	0,14019	0,02597		0,00013	
2-butanone (1)+1,1,2,2-tetrachloroethane (2)						
$V_E, \text{cm}_3.\text{mol}_-1$	-4,40442	-0,26605	-0,26976	-0,09947	0,00032	
$R_E, \text{cm}_3.\text{mol}_-1$	-0,13382	0,34352	-0,05271		0,00017	



- 2-butanone (1)+1-chloropentane (2)
- 2-butanone (1)+1,4-dichlorobutane (2)
- 2-butanone (1)+1,3-dichloropropane (2)
- 2-butanone (1)+1,1,1-trichloroethane (2)
- 2-butanone (1)+1,1,2,2-tetrachloroethane (2)



- 2-butanone (1)+1-chloropentane (2)
- 2-butanone (1)+1,4-dichlorobutane (2)
- 2-butanone (1)+1,3-dichloropropane (2)
- 2-butanone (1)+1,1,1-trichloroethane (2)
- 2-butanone (1)+1,1,2,2-tetrachloroethane (2)

(Tasic et al. 1992)

References

- Blanc, C., Elgue, J., & Lallemand, F. "MDEA process selects H2S". *Hydrocarbon Processing*, 1981.
- Canjar, L. & Manning, F. *Thermodynamic Properties and Reduced Correlations for Gases*, Gulf Publishing Company, 1967.
- Elgue, J., Peytavy, J., & Tournier-Lasserre J. "Recent industrial developments in natural gas sweetening by MDEA". Paper presented at the 18th World Gas Conference in Berlin, 1991.
- Gas Research Institute. *Thermodynamic Software Model (GRI – GLYCalc)*.
- GPSA. *GPSA Engineering Data Book, 1994 Edition*. 1994: 20-25.
- Katz, D., et al. *Handbook Of Natural Gas Engineering*. New York: McGraw-Hill Book Company, 1959.
- Kirk, R. & Othmer, D. (eds.). *Encyclopedia Of Chemical Technology, Volume 7*. New York: Interscience Encyclopedia, Inc., 1951.
- Maddox, R. *Gas And Liquid Sweetening*, 2nd Edition. Norman, OK: Campbell Petroleum Series, 1974.
- Mullins, B. et al. *Atmospheric Emissions Survey Of The Sour Gas Processing Industry*, EPA-450/3-75-076, U. S. Environmental Protection Agency, Research Triangle Park, NC: 1975.
- NACE MR0175. Sour Service. (2002) 08 February 2004. <<http://www.unitedvalve/com.nace.htm>>

- Royan, T. & Wichert, E., Options for Small Scale Sulphur Recovery. Paper presented at the CGPA Meeting. Calgary: 1995.
- Selleck, F. et al. "Phase Behaviour in the Hydrogen Sulfide – Water System". *Industrial and Engineering Chemistry* 44 1952: 2219-26.
- Sharma, S. & Campbell, J. "Predict Natural-gas Water Content with Total Gas Usage". *Oil and Gas Journal*. 1969.
- Song, K. & Kobayashi, R. Water Content Values of a CO₂ – 5.32 Mol Percent Methane Mixture, Research Report RR-20, Gas Processors Association, Tulsa: 1989.
- . The Water Content of CO₂ – Rich Fluids in Equilibrium with Liquid Water or Hydrates, research Report RR-80, Gas Processors Association, Tulsa: 1984.
- Tasic, A. et al. *Journal of Chemical Engineering Data* 37 1992.
- U.S. Environmental Protection Agency. *Protocol For Equipment Leak Emission Estimates*, EPA-453/R-95-017: 1995
- . *Control Techniques For Hydrocarbon And Organic Solvent Emissions From Stationary Sources*, AP-68. Research Triangle Park, NC: 1970.
- West, J. "Thermodynamic Properties of Hydrogen Sulfide". *Chemical Engineering Progress* 44 1948: 287-92.
- Wiebe, R. & Gaddy, V. "Vapor Phase composition of Carbon Dioxide – Water Mixtures at Various Temps. And Pres. to 700 Atmospheres". *Journal of American Chemical Society* 1941: 475-77.

Section SSc

Social Science

Section coordinated by the Department of Social Science in the School of Doctoral Studies of the European Union

Head of the Department:

Professor Ingmar Bremer

Reviewers Body Coordinators:

Deputy Head of Department - Humanities:

Professor Anuar Shah

Deputy Head of Department – Social Sciences:

Professor Ivanna Petrovska

SSc Section Content

Article 1

Page 158

Foreign Aid and Economic Development

Author: Lawrence McMillan

Country: UK

Article 2

Page 166

Marcel Duchamp Artist's Works Profile and Analysis on the Nude in Art proposed by Duchamp and Cocteau

Author: Erika González-Ehrlich

Country: Mexico

Article 3

Page 179

The Effect of Islamic Azad University Activity on Branch of Trade Case Study: Firoozabad

Author: Mehrnoush Saadat Abadi, Nima Naddaf (MBA), Fateme Saeed Firouzabadi (BBA)

Country: Iran

Article 4

Page 184

Service Failure and Recovery: Comparison Between Health Care and Automobile Service Station

Author: Anupam Krishna, G.S. Dangayach, Rakesh Jain

Country: India

Foreign Aid and Economic Development

Lawrence McMillan

Master of Science in Economics by the European Business School; Researcher for the
Business Intelligence Research Centre of London
147 St. John Street, 2nd Floor, London, UK
[Email: l_mcmillan2008@yahoo.com](mailto:l_mcmillan2008@yahoo.com)

Abstract

This literature review focuses on research studies that correlate the effects of aid on the macroeconomic development of countries in the third world. I undertake the review to assess whether researchers have strongly established, based on empirical evidence, that there is any link, whether positive or negative, between aid and economic growth. The Introduction section provides a more detailed rationale for this review. In the subsequent sections, I provide a short description of the econometric models commonly used by researchers in establishing the relationship between aid and growth. I also discuss the body of research that I have looked into, their basic features, general conclusions and some key criticisms against them. I devote a separate section identifying the gaps I observed in the current research and finally conclude that although there is a wealth of research done on the impact of foreign aid on growth, these appear to be inconclusive due to methodological problems. **Key words:** Economics, Foreign Aid; Economic Development.

Since its birth right after World War II when the United States of America (USA) released billions of money to assist Europe (Sogge, 2002) in reconstructing the latter's economy, foreign aid has been assumed to directly induce or at least influence economic development in a recipient country. Most donor rhetoric perpetuates this association (World Bank, 1998). Many studies have been undertaken to try to assess if aid actually fulfils its main objective, that is, to promote macroeconomic development in developing countries. After half a century characterized by some serious changes in world economy and politics (i.e. breakdown of Communism, globalization, terrorism), the link between foreign aid and economic growth remains at the center of debates on aid effectiveness.

The purpose of this literature review is to survey the available body of research done on the effectiveness of foreign aid as an agent for economic development of recipient countries, most of which come from the third world countries. In particular, this review aims to assess how strong the empirical evidence is that aid has been an effective agent for economic growth. In so doing, I hope to be able to identify areas where future research is required.

In reviewing the literature, I will attempt to provide a critical analysis of relevant research papers and articles. I will try to assess the merits or shortcomings of the research papers without leaning towards a particular perspective, for example economic structuralism or economic internationalism. Although I cannot guarantee complete

objectivity, I will make an effort to be academic in summarizing, identifying issues and exposing weaknesses or gaps of each piece of literature.

Extant literature on foreign aid is immense, hence I choose to focus on empirical research correlating aid with economic growth. Even with this delimitation, there is a wealth of information available and I admit a comprehensive review is totally not possible. Rather, I have chosen materials that appear most frequently in the total body of aid-growth literature. However, literature that primarily aim to establish the effect of foreign aid on political systems, governance and human rights in the recipient countries are not discussed in this review.

As in any debate, there have been two general but contending schools of thought with regard to the effectiveness of aid in spurring economic growth in the third world. Critics of foreign aid argue that it has had no effect on and even hurts the third world economies (negative aid-growth correlation). On the other hand, Supporters of foreign aid espouse that on average or in most cases, it has been an effective development tool (positive aid-growth correlation). A recent research trend has however emerged, which I see as a third school of thought. Researchers of this school qualify that aid can either be effective or ineffective depending on certain donor conditions and country circumstances (conditional aid-growth correlation). I have organized this literature review around these three groupings, though I will briefly discuss the econometric

models commonly used by researchers in assessing the impact of aid on economic growth.

Econometric Models Used in the Literature

Following are the two prevalent models utilized by analysts of the aid-growth relationship. These models are econometric but can be expounded in simplistic terms, which I endeavour to do in this section.

The Two Gap Model

The Gap Model popularized by Chenery and Strout (1966) ages ago is still in use in projecting the macroeconomic impact of foreign aid. This model has two components hence it is also commonly referred to as the Two-Gap Model. The first component is the relationship between investment and growth, wherein the level of growth is assumed to be dependent on the level of investment. The second component is the relationship between savings, which is assumed as a critical factor for investment expansion, and growth. With this model, analysts are able to determine the necessary level of investment to achieve a desired level of economic growth. Gaps occur if the investment is below the desired level and these gaps can be ascribed as either a savings gap or as a foreign exchange (or trade) gap. If a country is unable to fill this gap through imports, exports or production, foreign aid inflows or foreign capital inflows are needed so that it can grow more rapidly than its internal resources would otherwise allow. Hence an inflow of foreign aid should move a country's economy upwards.

Of course this model is not without criticism, most of which are questioning the assumptions of the model. Harms and Lutz (2004) point out that the gap model assumes that investment is the only factor in increasing output, whereas there are other determinants of growth (i.e. education, research and development). They also point out that not all aid is invested by the recipient country. Aid, as is any type of money flow, is fungible. It can be used for any purpose, and hence, aid cannot be assumed to all go into investment. A recipient country will naturally use part of the aid money for its consumption (government expenditures other than capital outlay) and part for investment. Harms and Lutz also recognize that in reality, aid availability is an incentive for corrupt administrations to intentionally lower their domestic investment efforts so that they get a continuous

stream of aid money from donors. Despite these criticisms, Devarajan, et al. (2002, p. 17) defend the two gap model saying that "it is a transparent and flexible framework for examining, for a large number of countries, the aid requirements of achieving the poverty goal". Most of the World Bank's research studies on foreign aid and growth rely on the two gap model.

The Poverty Trap Model

The poverty trap model is actually more of a theoretical framework than an econometric one. The earliest poverty trap model was used by Nelson (1956). Unlike the gap model which sees foreign aid as a way to raise investment and thus influence growth, this model assumes that growth is hampered by poverty traps which can come from various factors like low production capacity, high population, weak savings. Regardless of the causes, poverty traps are seen to compromise growth. Foreign aid, which is a temporary injection of capital, is assumed to help the economy get out of the poverty trap and take-off towards growth. Nelson sums it up nicely when he says that "increases in income and capital achieved through funds obtained from abroad can help to free an economy from the low-level equilibrium trap" (p. 904).

Unlike the gap model which necessarily requires the continuous and incremental inflow of aid into a recipient country, the poverty trap model requires a one time infusion of aid to spur economic growth in developing countries. But like the gap model, this model too has its limitations. Harms and Lutz state that it takes more than an infusion of aid for a country to get out of poverty and achieve economic growth. They say that the role of good governance and private capital is downplayed in the poverty trap model and that aid at best only provides a brief cure to poverty.

Anti Aid Literature

Generally speaking, economists and researchers who contribute to the anti aid literature espouse that aid has no affect on growth and that it may actually undermine it.

As early as the 1950s, questions on the effect of aid on economic growth have abounded. Economists like Friedman (1958) and Bauer (1972) called for an end in aid, arguing that it is not a necessary requirement for the economic growth of a country. Both Friedman and Bauer assert that foreign assistance to governments is dangerous because it increases the power of the elite in the recipient

governments, leads to corruption and hinders economic growth. In particular, Bauer noted that aid discourages the growth of private sector investments, encourages public sector-led growth (since aid is in fact money added to government coffers) thereby limiting growth and inhibiting development.

Although Bauer has been a leading critic of foreign aid, his ideas are grounded on very little empirical research and this is the main critique against his published writings. However, the lack of empirical evidence in Bauer's was made up for by other economists in the 1970s. Griffin and Enos (1970) were among the first to publish empirical research questioning aid effectiveness. They found, through simple correlation that there is a negative relationship between aid and growth in twenty seven (27) countries.

The pessimistic view of the aid pervaded as economists observed the persistence of poverty in developing countries. Throughout his life, Bauer (1991, p. 45) maintained his stance that aid is bad for development, "because aid accrues to the government, increases the government's resources, patronage, and power in relation to the rest of society. The resulting politicisation of life enhances the hold of governments over their subjects and increases the stakes in the struggle for power. This result in turn encourages or even forces people to divert attention, energy, and resources from productive economic activities to concern with the outcome of political and administrative processes and decisions".

These criticisms of foreign aid dates back four decades, hence I initially thought that such arguments would have been passé. However, I found that younger economists and new research still echo the old arguments. Other studies that show the lack of connection between lack of connection between aid and economic growth include Mosley (1980), Mosley et. al. (1987), Dowling and Hiemenz (1982), Boone (1994) and Kanbur (2000). These researchers converge in saying that aid fails to induce growth. They likewise suggest a variety of reason why.

Like Bauer, these researchers claim that aid is most often misused or corrupted by recipient governments. Kanbur, in particular, reports that aid fails to induce growth because of corruption, poor administration, tying up of aid with precious resources in recipient countries and questionable aid allocation decisions by donors. Another argument that these researchers espouse is that aid is a disincentive for the private sector to invest or to improve productivity. They cite the "Dutch disease" where aid

induces the recipient country's currency to appreciate and weakens the profitability of the production of export goods. For example, food aid causes local farm output prices to decline, thus reducing local farmers' income. They also argue that aid flows can reduce both private savings and government savings since aid impacts on interest rates and on revenues. These researchers also surmise that aid can help keep corrupt governments with poor economic policies and management capabilities in power, hence deterring growth and development.

Not all of these researchers generalize that aid is bad, though. Mosley, et. al. admit that there have been some individual aid projects that achieved success. Still, he posits the idea of a "micro-macro paradox", that the success of some individual aid projects cannot make up for the negative overall impact of aid on growth and development. While these studies have been influential, there are doubts as to their validity and quality. For one, how they measured the effect of aid on growth is unclear. Hence the attribution of negative growth with aid is also questionable, especially since a country's poor economic performance may be caused by a host of economic, social and political factors.

Because Boone (1996) examines the effect of aid on a variety of macroeconomic variables and several development indicators, his study is one of the most cited proofs that there is no significant, positive influence of aid inflows on investment and growth in recipient countries. However, some critics accuse that Boone assumed only a simple linear relationship between aid and growth, ignored potential endogeneity or other factors that may have direct or indirect effects on growth, and used an unconventional set of co-regressors (Clemens et. al., 2004).

Boone, however did not directly estimate the impact of aid on growth but used indicators such as consumption and investment as variables. He also tested the effect of aid on private and government consumption. Boone used a lot of other variables (i.e. the black market premium, indirect taxes and the inflation tax, infant mortality, life expectancy, and primary schooling). The sample comprises 96 countries and data are averages within the 1971-90 period. Basically, Boone's study shows that aid has no significant effect on any of the indicators frequently used to justify aid programs. Aid does not increase public investment, it does not affect taxation in developing countries, it does not lower infant child mortality and it does not raise life expectancy. Aid does, however, increase public consumption which can be construed to show that governments with access to aid

money tend to spend more. Boone (p. 322) sums up his study nicely, saying that “aid does not promote economic development for two reasons: poverty is not caused by capital shortage, and it is not optimal for politicians to adjust distortionary policies when they receive aid flows.”

A recent and also notable contribution to the anti aid literature is the one done by Easterly (1999 and 2001). This study strengthened critics’ skepticism about the effect of aid on growth. This study implies that aid money has been wasted as the predicted impact of aid (using the “two-gap” model) on output growth does not match the actual performance of a large sample of developing countries. Easterly found that aid had very little and in some cases no impact at all on the performance of majority of the individual countries he examined. In Easterly’s book (2001), he presents figures for Zambia, where its actual growth performance falls way below the predicted aid-induced growth.

As visualized in Easterly’s graphs of the Zambian case, there is a striking discrepancy or “gap” between the predicted growth line and the actual growth line. This has provided the strongest argument against a positive aid-growth relationship. Evident in Easterly’s study is that somewhere along the dynamics of aid provision and aid utilization in Zambia, not all aid went into investment or that not all aid-led investment were translated into economic growth. Other country examples include Congo, Haiti, Papua New Guinea and Somalia, which have been observed as not growing despite receiving sizeable foreign assistance. A large percent of the population in Africa and South Asia also remain impoverished despite being recipients of huge aid flows.

Pro Aid Literature

Literature reviewed under this section carry more or less the tune that there is a positive correlation between aid and growth, though there are some country exemptions. Researchers under this category maintain that the claims of the anti-aid school of thought are only partially correct, that aid can spur growth but its effectiveness decreases as the level of aid infused into the economy decreases. In other words, aid has diminishing returns.

Some early studies like Papenek (1973) and Levy (1988) found that aid had a positive impact on growth, hence sparking the debate between among economists and researchers. These analysts believed that aid increases growth by augmenting savings, financing investments, and adding to the capital stock. They argue that aid also helps increase productivity, especially aid in health or education

programs. They also consider the transfer of knowledge and technology from rich countries to poor countries as a positive effect. Like the early anti-aid literature, these claims were barely substantiated with empirical research. That there is an absolute positive correlation between aid and growth was more a belief than an actual fact since research at this time was focused on testing a linear relationship between aid and growth.

Sometime in the 1990s, researchers tested the diminishing returns hypothesis, that aid does spur growth but only up to a certain extent. Such studies include Durbarry et al. (1998), Dalgaard and Hansen (2000), Hadjimichael, et. al. (1995), Hansen and Tarp (2000 and 2001), Lensink and White (2001) and Dalgaard, et.al. (2004). These studies say that although aid has no absolute positive relationship with growth, there is evidence that the higher aid flows are, the more rapid growth is. Stiglitz (2002) and Stern (2002) have likewise argued that aid may have failed in some cases but it has undoubtedly been supportive of growth in some countries and prevented decline in other countries.

In support of their thesis, most of the above researchers identified successful aid recipients like Korea, Taiwan, Indonesia, Uganda and Mozambique and to specific programs like the aid-financed campaign against river blindness and oral rehydration therapy. They also argue that ever since the concept of aid for growth and development in the third world was introduced after the second world war, there has been significant improvement in the poverty, health and education indicators in many countries.

However, I think that these claims (i.e the growth of countries and the improvement of poverty indicators are caused / partially caused by aid) still need to be substantiated. For one, the growth and development of a particular country cannot be directly attributed to the inflow of aid. For another, data on aid is most often hard to isolate, unless one is evaluating a specific project. The data on official development assistance (ODA) which economists most commonly use in measuring the impact of aid on growth is an aggregate and may not be representative of the actual aid money that is brought into a country. Economists are still in search of a measure that firmly and exclusively establishes the link between aid and growth. One analyst, Roodman (2003), conducted sensitivity analyses on some of these studies and only the modelling of Dalgaard, et.al. and Hansen and Tarp turned out to be robust.

I found that Hansen and Tarp (2000) provides a summary of about thirty (30) studies published between 1968 and 1998. Individually, these studies estimate the relationship between aid and investment growth, albeit

with less sophisticated econometric modelling methods and smaller data sets. Hansen and Tarp proceeded to aggregate the results of these studies, in other words, they conducted a meta-analysis. The results of the meta analysis strengthen the claim that foreign aid raises the level of investment in recipient countries and Hansen and Tarp's study is frequently referred to in the newer studies as well.

A newer study by Hansen and Tarp (2001) used newer data from fifty six (56) countries from 1974 to 1993. Similar to this study is an earlier one conducted by Feyzioglu et al. (1998) who used data from thirty eight (38) countries within the period from 1971 to 1990. Both assessed the effect of effect of aid on investment and both found a positive relationship.

Qualified View

Probably because the two camps of aid researchers above cannot generalize that aid has absolutely no effect on growth and vice versa, other researchers are taking a different approach. Instead of ascertaining whether aid has a positive or negative relationship with growth, they endeavor to see where and in what circumstances can foreign have a positive or negative impact on growth. I believe this is a more "enlightened" approach to the aid-growth analysis. As apparent from the anti-aid and pro-aid literature, the effectiveness or ineffectiveness of foreign aid in spurring growth is on a case to case basis. It is therefore rational to identify the key factors that cause aid to work or not work for growth. The studies under this category may be further grouped into those that identify country specific factors and those that point out donor specific characteristics that provide conducive environment for aid to spur growth.

A number of research point out that in most circumstances, foreign aid worked wonders in countries with good policies and strong institutions. One such study was done by Isham, Kaufmann, and Pritchett (1995). In an effort to find out which circumstances are conducive for aid-induced growth, Isham, Kaufmann and Pritchett, using World Bank data, were able to link good performance of aid-financed projects in a certain country to that country's policies on civil liberty. Somehow, they found that there is a statistical and possibly (they say) a causal link between these two variables. Apparently, aid projects in countries that practice the best civil liberties have a higher economic rate of return than those in countries with poorer civil liberty systems. Interestingly, this study found that the type of government (i.e. authoritarian, democratic) has no effect on aid project performance.

One of the most popular studies providing evidence that aid has a positive effect on growth depending on the political-economic environment is Burnside and Dollar (2000). In a nutshell, Burnside and Dollar stipulates that aid encourages growth in countries with good policies. In this case, a country has good policy environment if there is low inflation, low budget deficit, and no protectionism meaning trade is relatively open. Statistical regression is the main method used by Burnside and Dollar to arrive at a composite variable that reflects the above criteria for a good policy environment. Using standard growth regression, Burnside and Dollar compared the interaction of aid (in percent of gross national product) with the policy variable and concluded that "the impact of aid is greater in a good policy environment than in a poor policy environment" (p. 859) and further suggested that "making aid more systematically conditional on the quality of policies would likely increase its impact on developing country growth" (p. 864). This study gained favour from aid regime supporters because it explains why aid has supported growth in several countries (Korea, Botswana, Indonesia, Mozambique and Uganda) while at the same time not influencing growth in others (Haiti, Liberia, Zaire and the Philippines). Another study done by Dollar in collaboration with Collier (2001) further substantiates the Burnside and Dollar findings. They find that additional aid of about one percent of gross domestic product increases, on average, the rate of economic growth by about 0.6 percent in countries with good policies, 0.4 percent in countries with average policies, and 0.2 percent in countries with poor policies.

Hence, the Burnside and Dollar study has largely influenced donor policies on aid. Their findings are consistent with donor claims that individual aid projects do impact on growth positively. This study also explains the "macro-micro paradox". Hence, donors like the World Bank are drawing on this study in rethinking the way they allocate aid and assess aid effectiveness. The main thesis of one of World Bank's policy research reports in 1998, for example, maintains that "money matters in a good policy environment" (p.28). The implication of this trend in the aid regime is that donors, in the near future, are more likely to allocate more aid to countries with good policies although this is yet to be seen.

After the 2000 study of Burnside and Dollar, other researchers followed suit and studied the interaction between aid and other country specific variables that may possibly explain the conditional relationship between aid and growth. Collier and Dehn (2001) linked aid effectiveness with the actual occurrence of external shocks (i.e. export price

shocks, climatic shocks). Guillaumont and Chauvet (2002) studied the variable of economic vulnerability (i.e. terms of trade). Collier and Hoeffler (2002) looked at conflict (i.e. policy warfare, totalitarian government) while Dalgaard, et al. (2004) looked at geographical factors (i.e. country location). These studies, however, were not as popular as Burnside and Dollar since statistical tests conducted by Roodman (2003) showed that the relationship of the above variables with aid-induced growth is fragile and modest at best.

Observed Gaps in Existing Research

There is a wealth of information dealing with foreign aid and growth and as apparent from the literature I have reviewed, there is no consensus at all on whether aid can indeed fuel growth at the macroeconomic level. Statistical and economic analysts, although employing identical or at least similar models of analysis, have arrived at different conclusions. To some, the impact of foreign aid on macroeconomic growth is negative, to some it is positive, and to some it is dependent on country-specific conditions. Furthermore, the methods used in each study have been subjected to sensitivity analyses and have been found lacking in one way or another. Evidently, a method or model of analysis has yet to be perfected in assessing the aid-growth correlation. I believe this is an area where economists can further delve into. The major challenge here is to develop a model that is sensitive enough to capture all the nuances of the aid-growth relationship, one that more or less takes into consideration the possible factors that affect this relationship (not just savings and investment as in the two gap model). Or economists can finetune the existing ones to make their results more robust and hence less misleading.

I have also observed that in the studies I have reviewed, the time factor was rarely mentioned. Assuming that we accept the traditional growth theory that aid increases investment and investment produces growth, it will still take a lot of time before we see the impact of the former on the latter. Hence, in assessing aid and its effects on growth, the time lag needs to be factored in as a variable. Interestingly, most researchers are silent about this time factor in their analyses. Aside from the time factor, the treatment of aid in aggregate terms (researchers use macro ODA data) may not be ideal because there are different types of aid which form part of the aggregate data but are not actually directed towards growth. Hence, using muddled ODA data

in analysing the impact of aid on growth may also muddle the results. Researchers still need to find a way to isolate data on aid that is intended solely for economic growth.

The studies I have reviewed all do cross country regressions as they try to establish the aid-growth causality, yet come up with different conclusions. Most of them work on the same data sets over the same period of time (at least the recent studies), but still come up with conflicting results. The problem, I believe, is in the attempt to generalize the effectiveness of aid across all countries, an attempt to globalize the benefits of aid. In doing so, researchers neglect to consider that each country is differently situated, differently governed and differently structured. Hence the macro approach or generalizing that foreign aid has a positive (or negative) effect on growth should not be the way to go in doing research on the aid-growth relationship.

The recent research trend (the qualified views) which attempt to identify country specific factors that allow aid to be effective, I believe is a more reasonable approach to aid and economic growth. Most researchers are now delving into this matter at present but the country specific factors that have been identified that facilitate aid-induced growth still generate debate. More research is necessary to firmly establish the link between a good policy environment and aid-induced growth (Burnside and Dollar's thesis). Since this type of research looks into country specific characteristics, I believe that country case studies is the way to go.

While country specific factors are being studied, I think it is also wise to study donor specific characteristics that could have an influence in inhibiting or facilitating aid-induced growth. After all, the foreign aid system is an interplay between a country and a donor. I have not come across any study that does this and I believe this is where future research should be focused on. Donor policies, practices and systems in providing aid need to be examined thoroughly. Donor procedures in aid provision could be one of the reasons why aid has not induced growth in some developing countries.

Conclusion

In this review, I have shown that despite the vast literature linking foreign aid and economic growth, there is yet no conclusive evidence that these two are positively (or negatively) correlated. This is basically due to the weaknesses of the models being used to assess this relationship. Most of the pro-aid and anti-aid studies

outlined above use either the two-gap model or the poverty trap model and deal with more or less the same data sets but still arrive at different conclusions.

The basic fault in the studies, I believe, is the attempt to generalize the effect of aid on growth across all countries. Because each country has varying features, and growth is affected by other variables, which the models cannot possibly incorporate, most studies, both under the anti-aid and the proaid literature, have been found lacking in merit.

Recent research, however, are moving towards identifying country specific factors responsible for the success or failure of aid in promoting growth. I believe this research trend is on the right track and where future research should concentrate on. It would also be prudent to study donor specific characteristics that possibly facilitate or hinder aid effectiveness in promoting growth, an area where there is yet very little research on.

References

- Bauer, P. (1972). *Dissent on development*. Cambridge: Harvard University Press.
- Bauer, P. (1991). *The development frontier: essays in applied economics*. London: Harvester Wheatsheaf.
- Boone, P. (1994). "The impact of foreign aid on savings and growth". Center for Economic Performance Working Paper No. 677, London School of Economics.
- Boone, P. (1996). "Politics and the effectiveness of foreign aid". *European Economic Review* (40): 289-329.
- Burnside, C. and Dollar, D. (2000, September). "Aid, policies, and growth". *American Economic Review* (90): 847-868.
- Chenery, H. and Strout, A. (1966). "Foreign assistance and economic development". *American Economic Review* (56): 679-733.
- Clemens, M., Radelet, S. and Bhavnani, R. (2004, November). *Counting chickens when they hatch: the short-term effect of aid on growth*. Center for Global Development Working Paper 44.
- Collier, P. and Dehn, J. (2001). *Aid, shocks and growth*. Policy Research Working Paper No. 2608, Development Research Group, World Bank, Washington.
- Collier, P. and Dollar, D. (2002). "Aid allocation and poverty reduction". *European Economic Review* 45(1): 1-26.
- Collier, P. and Hoeffler, A. (2002). *Aid, Policy and Growth in Post-Conflict Countries*. Policy Research Working Paper No. 2902, Development Research Group, World Bank, Washington.
- Dollar, D and Collier, P. (2001). "Can the world cut poverty in half? How policy reform and effective aid can meet the international development goals". *World Development* 29(11): 1787-1802.
- Dalgaard, C. and Hansen, H. (2000). *On aid, growth, and good policies*. CREDIT Research Paper No. 00/17, Centre for Research in Economic Development and International Trade, University of Nottingham.
- Dalgaard, C., Hansen, H. and Tarp, F. (2004), "On the empirics of foreign aid and growth", *Economic Journal*, 114 (496): 191-216.
- Dowling, M. and Hiemenz, U. (1982). *Aid, savings, and growth in the Asian region*. Report No. 3, Economic Office, Asian Development Bank.
- Durbary, R., Gemmell, N. and Greenaway, D. (1998). *New evidence on the impact of foreign aid on economic growth*. CREDIT Research Paper No. 98/8, Centre for Research in Economic Development and International Trade, University of Nottingham.
- Easterly, W. (1999). "The ghost of financing gap: testing the growth model used in the international financial institutions". *Journal of Development Economics* 60 (24): 423-438.
- Easterly, W. (2001). *The elusive quest for growth: Economists adventures and misadventures in the tropics*. Cambridge: Massachusetts Institute of Technology Press.
- Friedman, M. (1958). "Foreign Economic Aid". *Yale Review*, 47 (4): 501-516.
- Feyzioglu, T., Swaroop, V. and Zhu, M. (1998). "A panel data analysis of the fungibility of foreign aid". *World Bank Economic Review* (12): 29-58.
- Griffin, K. and J. Enos. (1970). "Foreign assistance, objectives and consequences". *Economic Development and Cultural Change*, 18.
- Guillaumont, P. and Chauvet, L. (2001). "Aid and performance: a reassessment". *Journal of Development Studies* 37: 66-92.
- Hadjimichael, M., Ghura, D., Muhleisen, M., Nord, R. and Ucer, E. (1995). *Sub-Saharan Africa: growth, savings, and investment 1986-93*. Occasional Paper 118, International Monetary Fund, Washington.

- Hansen, H. and Tarp, F. (2000). "Aid effectiveness disputed". *Journal of International Development*, 12: 375-398.
- Hansen, H. and Tarp, F. (2001), "Aid and growth regressions", *Journal of Development Economics*, 64: 547-570.
- Harms, P. and Lutz, M. (2004, September). The macroeconomic effects of foreign aid. Discussion Paper no. 2004-11, Department of Economics, University of St. Gallen.
- Isham, J., Kaufmann, D. and Pritchett, L. (1995). Governance and returns on investment: an empirical investigation. Policy Research Working Paper Number 1550, World Bank, Washington.
- Kanbur, R. (2000). "Aid, conditionality and debt in Africa". *European Economic Review* 32(9): 409-22.
- Lensink, R. and White, H. (2001). "Are there negative returns to aid?". *Journal of Development Studies*, 37(6): 42-65.
- Levy, V. (1988). "Aid and growth in sub-Saharan Africa: the recent experience". *European Economic Review* 32(9): 1777-1795.
- Mosley, P. (1980). "Aid, savings, and growth revisited", *Oxford Bulletin of Economics and Statistics* 42(2): 79-96.
- Mosley, P., Hudson, J. and Horrell, S. (1987). "Aid, the public sector and the market in less developed countries". *Economic Journal* 97(387): 616-641.
- Nelson, R. (1956). A theory of the low-level equilibrium trap in underdeveloped economies. *American Economic Review* 46: 894-908.
- Papenek, G. (1973). "Aid, foreign private investment, savings, and growth in less developed countries". *Journal of Political Economy*, 81(1): 120-130.
- Radelet, S. (2003). Challenging foreign aid: a policymaker's guide to the millennium challenge account. Center for Global Development, Washington.
- Roodman, D. (2003). The anarchy of numbers aid, development and cross-country empirics. Centre for Global Development Working Paper No. 32.
- Stern, N. (2002). "Making the case for aid," in World Bank, A case for aid: building a consensus for development assistance. Washington: World Bank.
- Stiglitz, J. (2002, April 14). "Overseas Aid is Money Well Spent," *Financial Times*.
- Sogge, D. (2002). Give and take: what's the matter with foreign aid? USA: Palgrave MacMillan.
- World Bank. (1998). Assessing aid: what works, what doesn't and why. USA: Oxford University Press.

MARCEL DUCHAMP

Artist's Works Profile and Analysis on the Nude in Art Proposed by Duchamp and Cocteau

Erika González-Ehrlich

Art student and candidate to Master of Arts at the School
of Doctoral Studies of the European Union in Brussels, Belgium

Email: ege.arts@sds.eu

Abstract

Marcel Duchamp (1887-1968) was one of four siblings who became artists in the period of intellectual and artistic ferment that saw out the last decades of the old century and extended beyond World War I. Duchamp's early interest was in painting and Cubism and much of his most influential work was related to Dada practice. But Duchamp was ultimately the most independent of artists--eventually becoming independent of art itself. Much of his influence derived from gestures or positions related to the nature of art, and a great deal of his fame rests on works consisting of ordinary objects altered or 'readymade.' But Duchamp's masterpiece is usually held to be the glass, metal, and paint construction entitled The Bride Stripped Bare by Her Bachelors, Even (1915-23), frequently known simply as Large Glass. On the other hand the influence of Marcel Duchamp and Jean Cocteau on the use of the nude as a subject in art falls outside the usual categories of influence. Duchamp's singular experience with his Nude Descending a Staircase (1912) was instrumental in his decision to turn from any conventional type of art career and become the twentieth-century archetype of the anti-artist. Even though the popular 'scandal' surrounding the painting was of a type that would have launched a more typical artistic career, and his intentions with regard to the Nude were somewhat conventional, Duchamp's subsequent path was devoted to the exposure of art's ways and means. The nude, however, was one subject whose dimensions fascinated him and at the end of his career he created the disturbing multimedia work Given: 1. The Waterfall 2. The Illuminating Gas (1946-66) which, in some senses, was a definitive response to the tradition of the nude in art. Cocteau, from another perspective, generated little scandal with his frankly homoerotic nudes--largely in a style derived from Picasso's classicizing period--because they were, for the most part, available only to a very limited audience. Cocteau was primarily a writer and filmmaker and his drawings constituted only a very small portion of his artistic output. But his style of drawing was very popular and in the work that was more widely available his nudes of both sexes frequently possess a frank, playfully erotic charge. Neither of these artists created a great number of nudes but the work they did contributed, in very different ways, to the rethinking of the nude as subject. **Key words:** Marcel Duchamp, Dada Art, Contemporary Art, Nude in Art, Jean Cocteau.

Duchamp's Profile and Works

This Large Glass single piece, left unfinished according to Duchamp, was the subject of numerous drawings, sketches, published notes, and etchings, and occupied a significant portion of the artist's active career. It is a Dada piece in terms of its emphasis on machine imagery, chance, jokes and puns, mysterious allusions, and eroticism, as well as its notional relationship to science and mathematics. All these elements display Duchamp's connections with the so-called New York Dada movement. But the Large Glass was a work conceived apart from important Dada considerations--such as spontaneity and the rejection of reason--and it contained the germ of some of Duchamp's future ideas. Yet, like the work of all the various art movements of the 1890-1920 period, Duchamp's singular achievement in Large Glass

was the product of ideas that grew out of the political and social restlessness of the era and from the ideas developed by its writers and philosophers.

The European world had changed radically in the course of the nineteenth century with the triumph of the Industrial Revolution and the rise of new European empires. Official art and culture, largely seen as the servant of the ever-expanding bourgeoisie, "had served conservative political ends," especially in France where government sponsorship of academic art traditions led, in part, to the link between the artistic and political radicalism that opposed the old order.¹

The institutionalized hypocrisy and anti-Semitism revealed in the Dreyfus Affair, which began in 1894, was highly divisive and, though not all avant-garde artists supported Dreyfus, the affair "reinforced the perception

of artists and intellectuals as defenders of ideals and moral positions irrespective of conventions."^{III} The growth of avant-garde movements throughout Europe--the various German Expressionists, the Italian Futurists, the Vienna Secession--coincided with the emergence of many new ideas in politics, philosophy, and psychology. Intensifying urbanization, the mechanization of society, and the rapid development of scientific and technical innovations produced a feeling of unstoppable change whose eventual destination was unknown but often wildly imagined as possessing possibilities that were the cause of increasing uneasiness. The materialism of industrialized society seemed to be displacing older values and the social order lacked the stability of former times. In politics Karl Marx's and Friedrich Engels' Communism and the Anarchism of Mikhail Bakunin and Petr Kropotkin held that the progressive nature of European capitalism was an illusion. In literature, writers such as the playwrights Henrik Ibsen and Oscar Wilde, and the novelist Fyodor Dostoevsky "exposed the social and moral dilemmas associated with materialism," while the Symbolist poets in France began to reject the notion of a realist basis for art and promoted the "world of the imagination [as] a source of spiritual renewal."^{IIII}

Among philosophers the most influential were Henri Bergson and Friedrich Nietzsche. The latter's ideas were based on Arthur Schopenhauer's metaphysical ideas, i.e., "that the truth of things lay behind surface reality."^{IV} Nietzsche's metaphysics served as a counterweight to the increasingly prevalent, materialist-oriented positivism that argued for the discovery of truth via observation and experimentation. Ibsen, August Strindberg and, in music, Richard Wagner were heavily influenced by Nietzsche and these artists conveyed their thought "through the experiences of isolated, frustrated, or doomed individuals."^V Nietzsche's philosophical autobiography *Ecce Homo*, which presented his ideas in digest form, was popular reading among artists and after the turn of the century painters as diverse as Picasso, deep in his Blue Period, the Fauves, and the new generation of German Expressionists in Die Brücke were producing art with a new type of content. Their works were more definitively urban in character and reflected the generalized anxiety that had begun to characterize life in the industrialized world. The roots of this new sensibility can be traced back to Nietzsche's ideas as well as to general social changes that had taken place in the preceding generations. The new sensibility's sources lay in, the realization of the appalling emptiness of a world in which God, as Nietzsche had declared, was dead; in less clandestine but no less

tormented sexuality; and in the pressures of mass society in which the individual discovers and maintains his integrity only with difficulty. The artistic response to these spiritual events is the often frantic search for self-expression.^{VI}

In France, Bergson had, by 1889, reached the conclusion that the only source of true knowledge was "intuitive experience" and that "art was a direct revelation of such experience."^{VII} Since mechanical means of representation--photography, motion pictures, sound recordings--were becoming increasingly important, the role of the arts as representational media seemed to be far less important. Bergson's lectures in the early 1900s also dealt with the notion that "the nature of experience was in a constant state of flux" and the individual's perceptions were, rather than an ordered, rational series of chosen moments, "a multiplicity of perceptions and memories" in which both the conscious and the unconscious mind played their parts.^{VIII} Bergsonian instability became a subject for art as painters, some of whom attended Bergson's lectures, broke up forms into series of impressions that generated the Cubist movement.

The psychoanalytic revolution initiated by Sigmund Freud also produced a new view of human behavior centering around the idea that "behind all rational façades lurked suppressed insecurities and desires determining every response."^{IX} Freud's ideas, many believed, exposed the sham of bourgeois ideals and morals, much as political theorists had done in regard to bourgeois notions of justice and individual freedom. There was also a wide interest among artists in a variety of spiritual and mystical systems, such as those of Rudolph Steiner and the Theosophists, which "encouraged aspirations toward a morally and socially responsible art," while stressing its innately spiritual nature.^X

All the stresses and strains of life in industrialized Europe and the variety of intellectual responses to new social forms and possibilities led to numerous avant-garde movements in the arts. Cubism was one of the most influential of these and it was here that the four young Duchamps began their careers.

Marcel Duchamp was the third of six surviving children born to a bourgeois Norman family. His father was a notary and the family lived very comfortably "while the elder Duchamp accumulated a tidy fortune" from his lucrative profession.^{XI} The children's maternal grandfather had been quite wealthy as well as having a serious interest in art. He too accumulated a fortune, as a shipping agent, before he devoted the remainder of his life to engraving and painting. The six children of this cultivated family were born in three groups of two. The eldest were Gaston (1875-1963) and

Raymond (1876-1918), and Marcel and his sister Suzanne (1889-1963) made up the second pair. Two sisters followed in 1895 and 1898.

The family's interest in art matured in both elder brothers who, after starting on careers in law and medicine, changed course and became artists. Gaston changed his name to Jacques Villon when he became a painter and engraver, and Raymond changed his last name to Duchamp-Villon when he became a sculptor. The precise reason for the name changes is not known, but, as Seigel speculates, it is likely that the family's great respectability might not have countenanced the sons' associates and the "sometimes immodest and risqué publications in which Gaston's early work appeared."^{xii} But if the first two children had some difficulty deciding on their course in life, for Marcel and Suzanne there seems to have been no problem, "with neither seriously considering any other kind of life."^{xiii} And, wholly respectable or not, the children received the support of their father throughout his life. Until he died, in 1925, he continued to provide financial support to all his artist children regularly, while, "for fairness' sake, carefully subtracting each one's advances from his or her share in the inheritance."^{xiv}

Thus Marcel Duchamp's entry into the world of art featured none of the struggle against class expectations that faced so many children of the bourgeoisie. His comfortable circumstances, Seigel suggests, contributed to his lifelong ability to take everything in stride and the fact that he never seemed "to need the reassurance that came from belonging to a group."^{xv} Duchamp's repeated decisions to go his own way relied, of course, on this kind of self-confidence but, more often than not, the nature of his plans made it necessary for him to be a solitary voice. Some ideas, such as readymades, do not merit too many repetitions and lose their impact in a crowd.

Duchamp's early work was, however, influenced by the predominant modes of the time. His earliest work shows the influence of the Fauves. But his elder brothers had become immersed in the Cubist initiative. Their Puteaux studios were the site of "regular discussions" with, among others, the painters Albert Gleizes and Jean Metzinger, the authors of Cubism's major theoretical work Du Cubisme (1912).^{xvi} Both Villon and Duchamp-Villon became successful artists. Villon had been a commercial illustrator, but his subsequent interest in Cubism was to endure. His paintings of the period 1912-22 earned him a sound reputation and, working from a strong "mathematical bias," Villon's Cubist-based works of the early 1920s were among the earliest non-objective works in French painting.

His career declined in the 1920s but he reemerged as a respected painter with "an attenuated Cubist manner" after World War II.^{xvii} Raymond Duchamp-Villon might have become a more important artist and was clearly developing a vigorous and inventive style when he died in the war. As Hamilton notes, Duchamp-Villon appeared to develop his approach in a manner similar to that of Marcel. The early works were less interesting in themselves "than as a series of steps by which he recapitulated the development of contemporary art and arrived at a startling reformation of the sculptural image."^{xviii}

In the autumn and winter of 1911-12 Marcel Duchamp worked in the Cubist mode "and appears to have absorbed fully the dictates" of the movement.^{xix} But Duchamp's most ambitious work in the Cubist style failed to please. His Nude Descending a Staircase (No. 2) (1912) was an attempt to analyze movement as well as form and it constituted such a marked departure from the prevailing style that a number of the older Cubists objected and he withdrew the painting from the Salon des Indépendants of 1912. Duchamp exhibited the work at another salon and at a Cubist show in Barcelona. But he made his most significant decision when he opted to send the work to the Armory Show in New York in 1913. The painting attracted so much attention that it rapidly became one of the best known works of art in the world--usually as an object of scorn and amazement.

Disgusted with the Cubists' reception of the work Duchamp withdrew to Munich for two months and worked on, among other things, a drawing entitled The Bride Stripped Bare by the Bachelors (1912). The painting reveals Duchamp's feelings of anger and betrayal which are "directly, even naively portrayed" in the scene in which the central figure of the 'bride' is attacked by figures on either side.^{xx} Inspired, perhaps, by the intensity of the moment Duchamp, who later, exaggeratedly, claimed the Munich trip had resulted in his "complete liberation," did accomplish a break with his previous interests and with the themes of the Cubists.^{xxi} He left behind his interest in linear motion and the dissolution of formal structures. This signaled the abandonment of art that was, in Duchamp's view, "retinal" in nature, i.e., art that "sought to communicate through the eye, not the mind."^{xxii}

Throughout 1912 Duchamp continued to produce paintings and sketches that concentrated on the theme of the 'bride' and the passage from virgin to bride. These works, Virgin, I (1912), Virgin, II (1912), The Bride (1912), and The Passage from Virgin to Bride (1912) worked through a number of important ideas and produced the

form of the bride herself as she was to be configured in the Large Glass. On his return to France Duchamp withdrew from the art world and took a position as a librarian in the Bibliothèque Sainte-Geneviève. The work was very easy and afforded Duchamp the opportunity to read in geometry and perspective--two subjects that were developing great importance in relation to the Bride project. In 1913 Duchamp, who had originally planned the great work as a canvas, painted Glider Containing a Water-Mill in Neighboring Metals, an imaginary machine that would reappear in the Large Glass, on glass and began to think of the future piece in this form. Also in 1913 Duchamp drew "his first full-scale study for the project" on the wall of a rented studio and had written many of the notes that he would later publish.^{xxxiii} The war, however, eventually drove him to seek refuge in America, following the famous path prepared by his ill-fated Nude.

Duchamp's Paris friend, the painter Francis Picabia, had also submitted his "superficially cubified" entries to the Armory Show and they "were almost as much remarked" as Duchamp's scandalous Nude.^{xxxiv} Early in 1915 Picabia, en route to his assignment in Cuba, deserted the French military and stayed on in New York where he "set about fashioning a metropolitan network of dealers and artists."^{xxxv} Duchamp followed a few months later and the two men rapidly became the center of "a vital mix of international talent and American patronage [that] fashioned itself into a comfortable avant-garde."^{xxxvi} Duchamp's principal patrons were Walter and Louise Arensberg, who eventually assembled the largest collection of his work. With their financial backing Duchamp did not do a great deal of painting in New York. He had begun to formulate ideas for the Large Glass in 1912 and by 1915 he had begun to realize the piece. By 1918 he painted his last canvas, a commission for the painter-collector Katherine Dreier entitled Tu M'--and abbreviated version of the expression "tu m'emmerdes," or "you bore the shit out of me."^{xxxvii}

Duchamp had definitively finished with painting and much of his dissatisfaction with his last work lay in the fact that it "relied upon ideas he had already developed in other works and therefore necessitated a certain degree of repetition" a practice he now detested.^{xxxviii} But Duchamp had also become interested in other approaches to art. He was privately working on the Large Glass and in 1916 he had begun to present his readymades. These works--ordinary machine-made objects such as shovels, bottle racks, and urinals that were 'selected' and signed by the artist--made it clear that "the idea of 'Art' was produced contextually," i.e., it is art because the artist says it is art.^{xxxix} Even more

importantly for his concurrent practice in the Large Glass, the readymades were often elaborately (if sometimes nonsensically and/or punningly) titled. The inclusion of language pointed up Duchamp's belief that "meaning itself is actively produced by viewing and speaking subjects, interpellated as they are by a variety of institutional positions."^{xxx}

But, important as Duchamp's challenges to the established order were, the goals of what came to be called New York Dada were very different from those of Zurich Dada, founded by Tristan Tzara and others. Although Duchamp, Picabia, and American artists such as Morton Schamberg and Man Ray "flouted artistic and social conventions, explored the possibilities of machine imagery, questioned the fixity of gender identities, [and] experimented with photography and assemblage" their approach was basically free of the traumatic associations that were a feature of Dada in Europe where the war continued to rage.^{xxxi} Duchamp and the others adopted the Dada label. But the Zurich group was witnessing "the violent collapse of an entire political and cultural order, while [the New Yorkers were] in possession of the security and the means with which to imagine a new order--or a happy state of disorder" and taunted the world but made no real effort to change it.^{xxxii}

Duchamp himself said that he had approved of Dada because he thought it a hopeful sign, but he could not consider himself a true Dada artist because, "I wished to show man the limited place of his reason, but Dada wanted to substitute unreason."^{xxxiii} Ultimately the "intentionally irrational Dada gesture" was a far cry from what Duchamp hoped to accomplish in his cherished Large Glass project.^{xxxiv} Painting had proved too limited and Dada gestures were too irrational for Duchamp's project. He believed that "art should exercise the intellect rather than simply indulge the eye" and he found the means to develop this idea in his revolutionary work the Large Glass.^{xxxv}

The Bride Stripped Bare by Her Bachelors, Even is oil paint and lead wire on glass and measures 225.5 x 175.6 cm, or 109 1/4 x 69 1/8 inches. Duchamp began planning the Large Glass in 1912 and took sketches and notes with him to New York in 1915. He then worked on the piece at intervals between 1915 and 1923, when he "abandoned it as definitively unfinished."^{xxxvi} The Arensbergs were the first owners of the work but, on moving to California in 1921, sold it to Katherine Dreier because it was believed to be too fragile to move. In 1927 the Large Glass was shattered while being returned from a New York exhibition. This was not known for a few years, however, as it remained

crated up. When the package was opened and the damage discovered in 1937, Duchamp repaired it, "piecing the fragments together with infinite patience and securing them between two heavier panes of glass, the whole bound in a new metal frame."^{xxxvii} The cracks in the two original panes of glass run in complementary directions because they had been packed one above the other when the accident happened. Duchamp eventually became quite pleased with the lines created by the cracks and declared that the work was finally completed "by chance" 14 years after he had supposedly abandoned it.^{xxxviii}

But he had never really abandoned the work at all. The Large Glass was seldom seen after the repairs until in 1952 Dreier's bequest sent it to the Philadelphia Museum of Art where the Arensberg collection was housed. Despite its absence from public view, however, it was well known to many artists "who respected it, if only by hearsay, as one of the most problematical works of modern times."^{xxxix} In 1934 André Breton wrote the first important essay on the work, bringing it to the attention of a broader public, "even though he had not seen it."^{xl} And, despite its invisibility in its original form, Duchamp himself kept its central imagery in the public eye.

If there were any need to define the distinctly non-Dada characteristics of the Large Glass, the conceptual nature of the work and its location at the center of a web of paintings, sketches, engravings, writings and other works would offer sufficient evidence of this when compared with the spontaneity so highly valued by Zurich Dada. This network of preparation and explanation preceded and followed the work. The claim that he had abandoned it in 1923 was followed, of course, by the publication of the notes in The Green Box in 1934. Via these notes, Duchamp was able "to provide an intrinsic, even if retroactively introduced, part of the structure of the work" which, as he claimed, was predicated on the idea of delay.^{xli} The idea of delay rested on Duchamp's location of the Large Glass' "action in a moment that never arrived" and it also described the work's "relationship to representation and meaning," but, in the long run, it also turned out to be a genuine temporal delay which was to extend much longer than anyone expected.^{xlii}

There was also the reconstruction of the broken work in 1937, and the construction of the various Boîtes en valise, or boxes in a suitcase, which were "small, portable collections that gather[ed] together nearly all the work he had done by 1935."^{xliii} This project, which came in a deluxe edition (of 20) with hand-colored "original reproductions" by Duchamp and an ordinary edition of 300, featured the Large Glass imagery quite prominently

and occupied Duchamp on and off from 1935 until his death.^{xliv} There was also a 1959 drawing entitled Cols alités in which the subject matter of the Large Glass "has been removed to an arid mountain landscape [where] the invisible waterfall that activates its water wheel," which set the imagined actions shown in the Large Glass in motion, "has evidently dried up" and "the erotic machinery of the Glass is faltering."^{xlv} This work was followed by a series of nine etchings that Duchamp made in 1965-66 for the first volume of Arturo Schwartz's The Large Glass and Other Works. In these etchings the principal elements of the work, the Nine Malic Molds, the Glider, the Seven Sieves, the Chocolate Grinder, the Oculist Witnesses, the Bride, and her Cinematic Blossoming, as well as the whole work as it was left in 1923 were replicated in three states each. The ninth in the series was a colored single-state etching entitled The Large Glass Completed (1965-66). As Wohl points out, the drawing Cols alités shows the machinery of the Large Glass in a corroding, neglected state but "still intact."^{xlvi} In the etchings, however, the work "was taken apart, reassembled, and even completed" very near the end of Duchamp's life.^{xlvii}

All of these works showed clearly that the Large Glass was not a dead issue for Duchamp. But the great surprise came after his death when it was revealed that he had spent the last 20 years working in secret, and in his typically desultory fashion, on his final work, Given: 1. The Waterfall 2. The Illuminating Gas (1946-66), commonly known as Given. In this multimedia work (encompassing found objects, electric light, tinted photography, and a representational, molded relief) the subject derived from "the hypothetical givens in the Preface note for the Large Glass," although the last work "did not proceed from notes or words, as the Large Glass had."^{xlviii} This surprising work, in which the viewer approaches a closed door and, peering through viewing holes, sees, as if through a hole in a brick wall, a naked woman, her head invisible, lying in front of a landscape with a waterfall and holding aloft a small lamp. The woman's hairless genitalia are clearly viewed between her spread legs and the whole has a certain "hyperreality, an excessive realism, which stages its eroticism as a 'too' obvious spectacle."^{xlix} Duchamp emphasized the hyper-visibility of the scene through the presence of the lighted lamp which is, of course, excessive in view of the very bright daylight in the scene. This trope "problematizes" the traditional equation of light and reason and the excessive illumination in the scene "makes us uncomfortable, breaking up the structure of voyeurism, its *raison d'être*--as the equation of sight and pleasure."^l Thus,

Duchamp presented a work that not only ceased to rely on the sort of wordy planning and explication surrounding the Large Glass, but also escaped any possibility of explanation because, except for three people who assisted him, the work was unknown to anyone at all until after Duchamp's death. He could offer no explanations. The last work was, then, the finished version of the Large Glass, which he had refused to produce throughout his life--keeping its imagery alive and subject to the "corrective function and [the] shap[ing of] the experience" of the work provided by his notes and repackagings.^{LI} It is Duchamp's "account of all the things he thought art should not be; it is the world of the Large Glass destroyed by being finished off."^{LII} Its hyperreality, lack of textual explication, and the absolute impossibility of explanation by the dead artist leave it as a piece of 'retinal' art of the type that Duchamp had rejected many years before when he finished with painting.

This desire to remove the work from the realm of verbal consideration was, as can be seen in retrospect, an essential part of Duchamp's strategy in sticking to his story of having abandoned art altogether. In his interviews given in the 1950s and 1960s Duchamp was always cagey about his position as a non-working artist. To Pierre Cabanne's question about the preservationist rather than 'making' role Duchamp seemed to have assumed in those years, Duchamp said this was correct because he "had already stopped making things."^{LIII} But when Cabanne asked, "Had you stopped absolutely?", Duchamp merely said, "Yes, but not absolutely. It had simply stopped, that's all."^{LIV} Thus, from the first dawning of the idea during his stay in Munich in 1912 until after he was dead, the ideas in the Large Glass truly were Duchamp's principal artistic concern.

A brief summary of the general scheme of the piece only touches on its complicated iconography. The work was essentially planned as the conclusion to a serial narrative begun in the Virgin, I (1912) and the other 1912 works. Its summary goal was an attempt "to embody in mechanized form the futility of human passion" and the combination of machine-based imagery and erotic content gives it an unusual tension that is, however, largely dependent on the viewer's knowledge of the artist's intentions.^{LV} Indeed, at one point in the process, Duchamp had planned that the viewer be in possession of his notes and able to reference them in conjunction with viewing the Large Glass, and this was again part of his plan when he published his notes in The Green Box.

There is a seemingly inexhaustible wealth of detail that can be elucidated for the various figures, machines, actions, and potential actions in the Large Glass. Their great

number tends to point up the fact that "these are games with an extremely elaborate but never fully specified set of rules, where precision and uncertainty combine to produce a world whose fluid energy never quite coalesces into stable forms."^{LVI} Despite the plethora of meanings inherent in the erotic narrative and the mechanization of human processes in the work, the rules of the game and Duchamp's own rules for their manufacture, i.e., his processes, are also central to the work's meanings. The processes employed by Duchamp were an intrinsic part of his primary aim in the Large Glass. Duchamp himself conveyed this idea in conversation with Breton in 1935, stressing the importance of novel processes that transcended those central to the art tradition. As Maharaj summarized these points, The Bride Stripped Bare by Her Bachelors, Even was shot through with one concern: to haul out painting and drawing from where they had become bogged down--the kind of spot where writing had come to a standstill before "Gutenberg." Duchamp sought to break with modes of seeing and representing that were tied to "painterly," manual processes, with intimacies of the handmade mark and its aura. The shift was toward a discipline of eye, hand, and mind that would enact something of high-speed mechanical techniques, their robotic jerks and swerves.^{LVII}

The terms employed by Duchamp are arbitrary in the sense that the objects represented in the Large Glass would seldom, of their own accord, evoke such terminology. They are also slightly arbitrary in the sense that Duchamp sometimes employed variant terms in The Green Box to describe the same objects or processes. Thus the notes he provided are, as Robert Lebel put it, "explicit except that they are often at variance or even contradict each other in details."^{LVIII}

The work consists of two horizontal rectangular panes, mounted one above the other. In the upper section the Bride, an organic-mechanical form taken from the painting The Bride (1912), floats at the left side of the pane. In the Large Glass the Bride lacks the fleshy coloring of the 1912 painting and, in Hamilton's opinion, this removes any sign that made her seem "remotely organic."^{LIX} Yet, in comparison with the more severely geometric forms in the lower, male, half of the composition, the organic quality of the Bride seems clear. The Bride is suspended below and to the left of her "halo" (or, as Duchamp sometimes called it, the "milky way" or the "cinematic blossoming"), the cloud-like shape, with three roughly square holes, located at the top center of the upper half. The openings in the halo reproduce the shapes achieved by pieces of gauze that Duchamp placed in front of an open window, allowing

them to be blown and shaped by the breeze. He then photographed the arrangement and used the breeze-blown shapes to create what he called "draft pistons." This was one of the instances in which Duchamp integrated chance into the work. Another such case was the eight small marks that appear below the halo (and a ninth, barely visible, inside it) which were created "by shooting matchsticks dipped in paint out of a toy pistol."^{LX} In terms of action, the Bride disrobes "as she simultaneously attracts and repulses her suitors," who are confined to the lower portion of the work.^{LXI}

In the lower half of the work, the Bachelors "orgasmic frustrations" are channeled, as "illuminating gas" through the tubes that connect them to the Bachelor Machine.^{LXII} The imagery in this section is far more complicated than in the Bride portion. The Bachelors are represented by nine shapes that Duchamp referred to as either "malic molds" or "uniforms and liveries." These shapes, which appear at the left, above the machines known as the "water mill" and "glider," are apparently based on illustration of men's clothing found in mail-order catalogues. They represent different male occupations: "priest, delivery boy, gendarme, cuirassier, policeman, undertaker, flunky, busboy, and stationmaster."^{LXIII} The nine figures are connected to the seven "sieves" or "parasols" that form a suspended arc over the so-called "chocolate grinder" at the center of the lower panel. The rods that connect the Bachelors with the sieves were shaped according to another of Duchamp's chance-based processes in which threads were dropped and the chance shapes they took were made into "standard" measures.

The pair of machines beneath the Bachelors consist of the water mill which sits inside the glider, sled, or chariot (three terms Duchamp used for this object). This section is reproduced almost exactly from Duchamp's first painting on glass, the Glider Containing a Water-Mill in Neighboring Metals (1913). Water was said to fall on the wheel, but the glider's movement was caused by bottle-shaped weights. At one point in the process Duchamp intended to have the glider powered by "the fall of brandy bottles, or lead weights shaped like brandy bottles," but, as wrote in his notes, he found this "much too far-fetched"--a sure indication that his notes were meant to include possibilities with which he only toyed, and should not all be taken as having significance for the ultimate meaning of the piece.^{LXIV}

The chocolate grinder is surmounted by a vertical rod called the "bayonet," which is topped by the two crossed, knobbed rods referred to as the "scissors." As the phallic design of the chocolate grinder makes clear, this portion of

the machinery represents the "principal male organ" in the piece.^{LXV} To the right are a small circle, over the scissors' blade, which was a magnifying glass, and, below the blade, three circular objects that seem to float. The larger shapes were taken from vision test chart and are called the "oculist witnesses." There were also a number of elements that were planned but were never included in the Large Glass. One is a small landscape that was to have been featured in interior of the glider, and which was somehow to have related "wheel and waterfall."^{LXVI} Others include the two visual effects that were to be featured above the magnifying glass; one was a "boxing match" and the other was "something called the 'Wilson-Lincoln effect,' an optical illusion that showed alternately one president or the other."^{LXVII}

The imagined operation of the Bachelor machine consisted of the molds filling with an "illuminating gas" which was, apparently, "a mode of male sexual energy," and when the "litanies of the chariot" were sounded the gas would expand.^{LXVIII} As it expanded it would rise from the molds, lose the shapes the molds had imposed on it, and move as "spangles" through the rods to the parasols. In the parasols the gas was converted to liquid form and would experience dizziness and spatial disorientation before falling, "in a corkscrew-like trajectory," to an orgasmic splash in the bottom right corner of the Bachelor space.^{LXIX}

The ramifications of the actions of the various elements in the Large Glass extend far beyond this brief summary. And, clearly, Duchamp developed an iconography that was uniquely his, "one that harbored meaning only within the narrowly established confines of an extremely personal, highly individualistic and self-reflexive narrative context."^{LXX} And within that context Duchamp was able to establish humor, irony, and meaning. Although it might be assumed that the invention of a wholly unique system of this sort was not absolutely necessary for carrying out his project of setting aside so-called 'retinal' art, it is important to note that by removing any trace of historical, traditional iconographic content Duchamp positioned the viewer entirely within his own system. This effect was, of course, reinforced by the publication of the notes which further constrained the viewer to look at the work through the lens of Duchamp's private system. And he not only created an iconography, he also developed a work process that was every bit as individualized.

Much of his inspiration for the idea of individualized processes seems to have derived from the example of the bizarre dramatic performances and fiction of the eccentric Raymond Roussel, whose Impressions of Africa Duchamp saw in 1912. Though Roussel did not expound on his method

until 1935 (and Duchamp did not know him personally), Duchamp may have "grasped or intuited" much of it in 1912 and certainly "had plenty of time to reflect on what he had seen and to read" Roussel's works, as he certainly did, afterward.^{LXXI} Duchamp later credited Roussel as the inspiration for the Large Glass and Seigel's description of Roussel's method could be applied almost in its entirety to Duchamp's own:

Roussel showed how the recourse to language games and mechanical imagery, which some interpreters have taken to signify the demise of personal subjectivity in art, actually allowed his works to become the scene for developing and acting out a series of highly personal themes and preoccupations. . . . his procedure served to disguise both what these obsessions were and how powerfully they moved him, casting them in a mold of seeming exteriority and objectivity. Much of the aura of mystery and incomprehensibility that surrounds his novels and plays arises from the way these two opposing currents, personal and impersonal, flow together in his work, each serving at once to highlight and to obscure elements of the other.^{LXXII}

Among the many strategic, inventive processes adopted by Duchamp for the Large Glass were the operations based on what he called "canned chance," the use of an individualized type of 'science,' and the investment of certain aspects of the erotic narrative with mathematical qualities.^{LXXIII} One of the best examples of his chance-based operations was the "standard stoppages" which were used, for instance, in the design of the rods connecting the Bachelors to the seven sieves. Duchamp described the process under the heading, "The Idea of the Fabrication," noting that "if a [straight horizontal] thread one meter long falls from a height of one meter straight on to a horizontal plane twisting as it pleases and creates a new image of the unit of length."^{LXXIV} Duchamp followed this process, allowing the string to fall onto canvas surfaces. He then took an impression of the string and affixed it to glass plates, from which he prepared wood templates "duplicating the subtle twists and curves of their chance configuration" and using the results as measuring tools.^{LXXV} He considered this elaborate process "a joke about the meter," the standard of measure which was based on "two scratches on a platinum-iridium bar housed in a temperature-controlled chamber in the Academy of Science."^{LXXVI} The measuring tool devised by Duchamp was, therefore, not only based on "canned chance" but was given a mockingly fixed status even though it was intended to be employed only for the private requirements of the artist.

In addition to the establishment of a new, random set of measurements, Duchamp's method also fulfilled his desire, as he said, "to strain the laws of physics just a little" and to this end he employed an ironic science devised by the playwright Alfred Jarry which he called "'pataphysics."^{LXXVII} This witty science was a science of the particular rather than the general; that is, individual exceptions ruled where in normal science generalizable rules were required. Duchamp used this type of science in the Large Glass in, for example, the operations of the glider which he said was "'emancipated horizontally' from gravity, so that it could slide without friction."^{LXXVIII}

The pleasant absurdities of his science were matched by the uses of mathematics in the Large Glass. His interest in genuine mathematical operations, such as the positing of a fourth dimension, was given a witty turn in, for example, imagining the fate of the individual who entered such a dimension. In the Large Glass the Bride is potentially susceptible to such a procedure and this is indicated by the halo or cinematic blossoming. As Adcock notes, blossoming is a term that can be applied to both sexual awakening and outward expansion. The cinematic blossoming refers, therefore, both to the Bride's imminent sexual initiation and her expansion into a fourth dimension. In mathematics, the "four-dimensional continuum opens and expands outward from normal three-dimensional space along axes not contained within three-dimensional space" and the Bride, who experiences the blossoming, "expands outward from a center into realms that are ninety degrees away from any direction that exists in normal space."^{LXXIX} Within that dimension Duchamp located various sexually charged possibilities--such as the transformation of gender and other complications that only total immersion in his system would render comprehensible.

The most remarkable fact about Duchamp's The Bride Stripped Bare by Her Bachelors, Even turns out to be its totality: it occupied him in one form or another from 1912 to 1968; it featured an iconography that was wholly personal and unique; this iconography was explicated by its author in a manner that left many holes and contradictions intact--requiring the viewer's immersion in the piece and its textual supplements for full comprehension (which might not be possible even then); its science and mathematics were equally recondite; and it was impossible to view the work in any fixed manner. Altogether Duchamp went as far in the Large Glass as he possibly could toward creating art that replaced the centuries-old tradition in which the purely visual seemed to dominate. Then, in his reversion to representation in Given, he supplied the illusion of distance

and space, the careful depiction of realistic detail, the color and compositional norms, the classic subject of the nude, and the elimination of intellectual content. In this work, however, the subject, the Bride Stripped, is presented with a literal quality that is as brutal in terms of the way the woman is depicted as Duchamp believed representational art was brutal in and of itself. The lightness and seriousness of the Large Glass are replaced by the brutish act of trying simply to recreate what already exists.

The Nude in Art, as Proposed by Duchamp and Cocteau

Duchamp's early work as a painter was heavily influenced by Cezanne, the Nabis, and the Fauves but in the spring of 1911, under the influence of his brother Jacques Villon, Duchamp began to work in the Cubist mode. He reduced his palette to a few somber colors and began to experiment with the Cubist notion of 'simultaneous aspects' that had been developed by Braque and Picasso. But instead of developing a single synthesized image from a number of aspects of the woman who was his subject Duchamp "represented the same figure on different planes and in different attitudes to express more of the individual as well as the concept of Woman" (Douglas Cooper 121). Duchamp was taken up by the painters Albert Gleizes and Jean Metzinger and had begun to be touted as a promising artist. But at the end of 1911 he had also begun to experiment with combining the different aspects of human figures "to represent a succession of evolving movements"-an idea that was being developed by the Futurists in Italy, whose manifesto had been published in Paris but whose work remained as yet unseen there (Douglas Cooper 121).

This work did not generate much enthusiasm among Cubist circles but Duchamp continued his experiments in his first version of Nude Descending a Staircase (1911-12). This figure still retains traces of its organic qualities and moves down a more or less realistic stairway leaving traces behind it as in a stroboscopic photograph. The figure is segmented (rather erratically) so that portions are truncated cylinders and it has a unity that makes it clear that it is a single figure. The second version of the Nude, however, eliminated any traces of the organic or realistic detail and showed successive stages of the movement so that, in effect, several figures could be said to move down the stairs. In Douglas Cooper's view Duchamp had, at this point, "turned the figure into a symbolic but seemingly powerful machine which rattles its metallic structure and devours the staircase as it descends" (124). The Nude is all planes and angles

with a few dotted lines signifying axial movement. It was not, however, what the Cubists had in mind and it was rejected by Gleizes and Metzinger who were the judges of the Paris Salon des Indépendants in 1912.

Duchamp did not protest and merely withdrew the picture though, as he said later, the experience "gave me a turn" and led to his decision to leave Paris for Munich (quoted in Seigel 61). The relationship of the painting to Cubism was ambiguous at best. The Futurists idea of representing movement in a picture struck the Cubists as far less interesting than their own idea of simultaneity in which the painter sought to capture "the coexistence of a number of separate ideas or mental experiences in a single instant" (Seigel 61). In Munich Duchamp began work on the organic-mechanical figure groups that were to lead to the development of the private iconography that was to be the basis for The Bride Stripped Bare by Her Bachelors, Even (1915-23), or the Large Glass, a work that occupied most of the rest of his life and that he never considered completed.

The Nude of 1912 went on to be exhibited in Barcelona and in Paris and provoked little interest. But it was finally selected, along with three other works by Duchamp, for the 1913 New York International Exhibition of Modern Art, better known as the Armory Show, where it proceeded to cause a sensation. The "explosion in a shingle factory," as it came, among other things, to be called drew crowds, inspired cartoons and parodies, and generally provoked the greatest modern art sensation of the early part of the century. Duchamp's works sold well and collectors, particularly Walter Arensberg, began to take an interest in him. The public's reaction to Duchamp's painting was, however, a curious blend of a desire to be shocked by the novelty of modern art and titillation provoked by the title of Duchamp's work. The reaction was based in part on the public's long-standing reaction, "a complex of jealousy and suspicion," toward "artists' special access to nude bodies" (Seigel 6). Although, for the most part, people paid lip-service to the notion that art raised "nudity out of the sphere of sensuality and into that of the ideal through the body's power to stand for formal perfection" this idea had already come under fire among artists themselves (Seigel 4). In his Olympia (1863) Manet had employed his gift for irony by delving into the area between the exaltation of the pure beauty of the human figure and the mere nakedness of his model--in part as a demonstration of the changes taking place in painting itself and in part as a demonstration of the curiously hypocritical stance that the academy took in works such as Alexandre Cabanel's Birth of Venus, also from 1863,

whose eroticism was undeniable but unacknowledged. The crowds endlessly massed around Duchamp's picture at the Armory Show in 1913 were as much titillated by the idea that the 'nude' had somehow escaped their perception as they were filled with the enjoyment of observing a work that, as Clement Greenberg said, "gave people enough clues to watch themselves being startled by the 'new'" (quoted in Seigel 6).

What the curious fame of this painting meant for Duchamp, whose success in America when he arrived in 1915 still surprised him, was the freedom to develop as he wished. The rejection of the Nude by orthodox Cubists had disturbed him more than he had let on. The fact that his fellow artists, including his two artist brothers, objected to the painting as not what they felt Cubism should be made him regard "such conservative and overtly dogmatic behavior [as] an aberration, particularly for artists who purport[ed] to be more open-minded than the general public" (Naumann 27). Duchamp had already decided to take up another profession (librarian) in order to be able to work freely at his art. The unexpected New York success meant that he was free to work in America. The result was not only "a radical break from his own earlier work, but [a] definitive break from the previously established and accepted conventions of the art-making process" (Naumann 26).

In subsequent work Duchamp was little concerned with representation per se, but in his last decades he returned to the nude and once again strove, as Manet had in 1863, to demonstrate the limitations of the accepted modes of art production via his treatment of the classic subject of the female nude. In the Given piece, a lighted, three-dimensional construction that is viewed through a hole in a door, Duchamp made specific reference to Courbet's notorious Origin of the World (1866) and presented the splayed thighs of a woman who lies in a realistically constructed field near a stream, holding a lamp in one upraised hand. Her face is not visible and the "seemingly aggressive and hostile treatment to which the female figure has been subjected" is quite disturbing (Seigel 113). But at the time of his early success, as Duchamp later remarked, he had begun to feel "obscured .crushed by the Nude" and he recognized the extent to which even with "techniques that dematerialized the body so that it seemed to have little to do with sexuality" there was a strong tendency to ascribe "immorality and decadence" to artists (quoted in Seigel 4, 6). By the latter part of his career Duchamp had become immersed in demonstrating the futility of the conventions of art-making and his Given returned to the subject of the

Nude in an effort to comment on both the representative in art and the eroticism that pervaded the nude subject--no matter how the artist approached it. The meticulous 'realism' of Given and the brutal display of the disembodied sexuality of the nude answer, in effect, the hordes who were so amused and so titillated by the 1912 painting.

Jean Cocteau's approach to the nude subject was certainly far less intellectual than Duchamp's but it was no less effective in focusing on the erotic nature of the representation of the human body. Cocteau's drawings display a remarkable consistency over the decades. His style seldom varied until near the end of his life when he began to design motifs for a small industry of ceramic wares and lithographs and even painted murals for a church. The style developed largely in conjunction with Picasso's development of a classicizing, linear approach to painting and drawing. Cocteau and Picasso were close friends and the writer persuaded the painter to accompany him to Italy during the First World War at the invitation of the Russian ballet impresario Serge Diaghilev. As Gertrude Stein noted, had it not been for Cocteau Picasso would never have gone; the painter "allowed others to make decisions, that is the way it is, it was enough that he should do his work" (quoted in Crosland 46).

Picasso's exposure to the classical treasures of Italy, however, had a profound effect on his work and he began to develop a simple, linear approach to somewhat colossal figures, simply attired in classicizing dress or nude, that was to characterize his work, on and off, for the next fifteen years. Cocteau's own drawings followed suit. The illustrations that now accompany his novel The Holy Terrors (Les enfants terribles) display the style as he used it in published work. The twinned profiles of his principal characters Paul and Elisabeth are shown, for example, on page 27. The pairs' shoulders are draped in indefinite pieces of cloth, like Republican busts, and their simply drawn hairstyles are also reminiscent of Roman sculpture. Their profiles feature strong noses and chins and eyes that look like the schematic carving of statues. They have a Roman gravity that is belied only by Elisabeth's eye which shifts to the side, giving her a suspicious air that appears to be directed at the viewer. In other drawings from this series, such as the depiction of the pair in the bath, the erotic charge between them is accomplished largely through the direction of their gaze and the long leg that each of them lifts from the bath in a crossed pattern that is at once comic and sexually vibrant (Cocteau 67).

Cocteau did not consider his graphic work to be merely illustrative and explained that the drawing most "formed

a different stage in the expression of an idea which had already appeared in some other form" (Crosland 149). The suite of drawings now published with *Les enfants terribles* (1929), for example, were actually a separate project that was published in 1934, and in 1950 Cocteau made a film of the novel as well. The classicized profiles and idealized bodies of the subjects of Cocteau's drawings received their most prominent expression, however, in works that either published, but for which Cocteau was "reluctant to acknowledge authorship" or were "for private circulation only" (Emmanuel Cooper 136, 139). Cocteau "present[ed] a respectable face to the public at large" but was openly homosexual within the circle of artists of whom he was and the American writer Gertrude Stein were the general leaders (Emmanuel Cooper 139). His efforts at combating social repression of homosexuality extended, therefore, to writing a work such as *Le livre blanche* (1928), a semi-autobiographical account of a young man's sexual history, but he published it anonymously. The authorship was an open secret among the artistic world however and when he published a second edition in 1930 it featured a number of explicitly homoerotic drawings by the author. These idealized nudes with "their handsome Greek-looking profiles, thick curly hair and well-built muscular bodies depict physical perfection" as conceived by Cocteau via classical examples (Emmanuel Cooper 136).

Even more explicit drawings were intended entirely for private circulation and the men in works such as the untitled drawing included by Emmanuel Cooper (137) with their large genitals, carefully outlined musculature, sexual postures, and dreamy, absorbed expressions represent an erotic ideal--almost a version of classical statuary come, vibrantly sexual, to life. The same ideal of male beauty expressed as a sexual ideal ever more explicitly as the works moved farther from the public eye. But the presentation of the homoerotic in the depiction of nudity--especially in nudes that so directly reflected the influence of the classical--was, in effect, an undermining of the general notion of the exalted, non erotic nude. If the art of the past supposedly lifted "nudity out of the sphere of sensuality and into that of the ideal" the art of Cocteau forced the viewer's attention to the erotic element by the singularity of his concentration of attention on the male figure--as opposed to the more traditional (since the Renaissance) female (Seigel 4). If, Cocteau's nudes imply, the viewer is inexorably reminded of the erotic focus of the artist's gaze even in the generalized (i.e., not sexually engaged) nudes then what is one to make of the female figures that have so effortlessly been passed off as the ideal in previous art?

In effect both Duchamp and Cocteau contributed strongly, albeit in very different ways, to the deconstruction of the nude-as-subject in art. Duchamp's 1912 painting inadvertently focused attention on the prurience of audiences and raised the question of whether the ideal representation of the nude was possible and his later work demonstrated his conviction that, even if it had been possible to present realistic yet idealized nudes, neither realism nor nudity was possible any longer in art. Cocteau, less deliberately, forced the thoughtful viewer to wonder what, exactly, was idealized in the traditional female nude. If the artist's focus was on the male rather than the female form and this generated audience discomfort or surprise then how could it be argued that there was no parallel idealization of the erotic in earlier nudes?

References

- Adcock, Craig. "Duchamp's Eroticism: A Mathematical Analysis." In *Marcel Duchamp: Artist of the Century*, ed. Rudolf Kuenzli and Francis M. Naumann, 149-67. Cambridge: MIT Press, 1987.
- Buskirk, Martha. "Thoroughly Modern Marcel." In *The Duchamp Effect*, ed. Martha Buskirk and Mignon Nixon, 190-203. Cambridge: MIT Press, 1996.
- Duchamp, Marcel. *The Writings of Marcel Duchamp*. Ed. Michael Sanouillet and Elmer Peterson. New York: Da Capo, 1989.
- Cocteau, Jean. *The Holy Terrors*. 1929. Trans. Rosamond Lehmann. New York: New Directions, 1957.
- Cooper, Douglas. *The Cubist Epoch*. 1970. London: Phaidon, 1994.
- Cooper, Emmanuel. *The Sexual Perspective: Homosexuality and Art in the Last Hundred Years in the West*. 2nd ed. London: Routledge, 1994.
- Crosland, Margaret. *Cocteau*. New York: Knopf, 1956.
- Gale, Matthew. *Dada and Surrealism*. London: Phaidon, 1997.
- Hamilton, George Heard. *Painting and Sculpture in Europe: 1880-1940*. Rev. ed. Harmondsworth, England: Penguin, 1972.
- Judovitz, Dalia. "Rendezvous with Marcel Duchamp: Given." In *Marcel Duchamp: Artist of the Century*, ed. Rudolf Kuenzli and Francis M. Naumann, 184-202. Cambridge: MIT Press, 1987.
- Maharaj, Sarat. "'A Monster of Veracity, a Crystalline Transubstantiation': Typotranslating the Green Box." In *The Duchamp Effect*, ed. Martha Buskirk and Mignon Nixon, 61-91. Cambridge: MIT Press, 1996.

- Molesworth, Helen. "Work Avoidance: The Everyday Life of Marcel Duchamp's Readymades." *Art Journal* 57 (1998): 50-63.
- Naumann, Francis M. "The Bachelor's Quest." *Art in America*, September 1993, 72-83.
- "Marcel Duchamp: A Reconciliation of Opposites." In *Marcel Duchamp: Artist of the Century*, ed. Rudolf Kuenzli and Francis M. Naumann, 20-40. Cambridge: MIT Press, 1987.
- Nesbit, Molly. "Last Words (Rilke, Wittgenstein) (Duchamp)." In *About Michael Baxandall*, ed. Adrian Rifkin, 84-102. London: Blackwell, 1999.
- Seigel, Jerrold. *The Private Worlds of Marcel Duchamp: Desire, Liberation, and the Self in Modern Culture*. Berkeley: University of California Press, 1995.
- Vetrocq, Marcia E. "The New York Pre-School." *Art in America*, June 1997, 82-87.
- Wohl, Hellmut. "Duchamp's Etchings of the Large Glass and The Lovers." In *Marcel Duchamp: Artist of the Century*, ed. Rudolf Kuenzli and Francis M. Naumann, 168-83. Cambridge: MIT Press, 1987.

Notes

- ⁱMatthew Gale, *Dada and Surrealism* (London: Phaidon, 1997), 11.
- ⁱⁱ*Ibid.*, 13.
- ⁱⁱⁱ*Ibid.*, 18.
- ^{iv}*Ibid.*, 19.
- ^vGeorge Heard Hamilton, *Painting and Sculpture in Europe: 1880-1940*, rev. ed. (Harmondsworth, England: Penguin, 1972), 141.
- ^{vi}*Ibid.*, 197.
- ^{vii}*Ibid.*, 76.
- ^{viii}Gale, 19.
- ^{ix}*Ibid.*, 26.
- ^x*Ibid.*, 28.
- ^{xi}Jerrold Seigel, *The Private Worlds of Marcel Duchamp: Desire, Liberation, and the Self in Modern Culture* (Berkeley: University of California Press, 1995), 16.
- ^{xii}*Ibid.*, 17.
- ^{xiii}*Ibid.*
- ^{xiv}*Ibid.*
- ^{xv}*Ibid.*
- ^{xvi}Gale, 85.
- ^{xvii}Hamilton, 262.
- ^{xviii}*Ibid.*, 274.
- ^{xix}Francis M. Naumann, "Marcel Duchamp: A Reconciliation of Opposites," in *Marcel Duchamp: Artist of the Century*, ed. Rudolf Kuenzli and Francis M. Naumann (Cambridge: MIT Press, 1987), 26.
- ^{xx}Seigel, 66.
- ^{xxi}*Ibid.*, 68.
- ^{xxii}*Ibid.*
- ^{xxiii}*Ibid.*, 87.
- ^{xxiv}Hamilton, 369.
- ^{xxv}Marcia E. Vetrocq, "The New York Pre-School," *Art in America*, June 1997, 82.
- ^{xxvi}*Ibid.*
- ^{xxvii}Francis M. Naumann, "The Bachelor's Quest," *Art in America*, September 1993, 74.
- ^{xxviii}*Ibid.*
- ^{xxix}Helen Molesworth, "Work Avoidance: The Everyday Life of Marcel Duchamp's Readymades," *Art Journal* 57 (1998): 50.
- ^{xxx}*Ibid.*
- ^{xxxi}Vetrocq, 82.
- ^{xxxii}*Ibid.*, 87.
- ^{xxxiii}Quoted in Hamilton, 376.
- ^{xxxiv}Hamilton, 376.
- ^{xxxv}Gale, 86.
- ^{xxxvi}Hamilton, 375.
- ^{xxxvii}*Ibid.*
- ^{xxxviii}*Ibid.*
- ^{xxxix}Hamilton, 375.
- ^{xl}Seigel, 89.
- ^{xli}Martha Buskirk, "Thoroughly Modern Marcel," in *The Duchamp Effect*, ed. Martha Buskirk and Mignon Nixon (Cambridge: MIT Press, 1996), 193.
- ^{xlii}Seigel, 96.
- ^{xliiii}*Ibid.*, 229.
- ^{xliv}*Ibid.*
- ^{xlv}Hellmut Wohl, "Duchamp's Etchings of the Large Glass and The Lovers," in *Marcel Duchamp: Artist of the Century*, ed. Rudolf Kuenzli and Francis M. Naumann (Cambridge: MIT Press, 1987), 168.
- ^{xlvi}*Ibid.*
- ^{xlvii}*Ibid.*
- ^{xlviii}Molly Nesbit, "Last Words (Rilke, Wittgenstein) (Duchamp)," in *About Michael Baxandall*, ed. Adrian Rifkin (London: Blackwell, 1999), 87.

^{XLIX}Dalia Judovitz, "Rendezvous with Marcel Duchamp: Given," in *Marcel Duchamp: Artist of the Century*, ed. Rudolf Kuenzli and Francis M. Naumann (Cambridge: MIT Press, 1987), 187.

^L*Ibid.*, 187-88.

^{LI}Buskirk, 192.

^{LII}Seigel, 113.

^{LIII}Quoted in Buskirk, 194.

^{LIV}*Ibid.*

^{LV}Gale, 87.

^{LVI}Seigel, 93.

^{LVII}Sarat Maharaj, "'A Monster of Veracity, a Crystalline Transubstantiation': Typotranslating the Green Box," in *The Duchamp Effect*, ed. Martha Buskirk and Mignon Nixon (Cambridge: MIT Press, 1996), 61.

^{LVIII}Quoted in Buskirk, 193.

^{LIX}Hamilton, 376.

^{LX}Seigel, 91.

^{LXI}Hamilton, 375.

^{LXII}*Ibid.*

^{LXIII}Seigel, 92.

^{LXIV}*Ibid.*, 89.

^{LXV}Hamilton, 375.

^{LXVI}Seigel, 92.

^{LXVII}*Ibid.*, 93.

^{LXVIII}*Ibid.*, 92.

^{LXIX}*Ibid.*

^{LXX}Naumann, "Marcel Duchamp: A Reconciliation of Opposites," 29.

^{LXXI}Seigel, 83.

^{LXXII}*Ibid.*, 84.

^{LXXIII}Marcel Duchamp, *The Writings of Marcel Duchamp*, ed. Michael Sanouillet and Elmer Peterson (New York: Da Capo, 1989), 33.

^{LXXIV}*Ibid.*, the unusual layout of the printed text of the notes reflects Duchamp's handwritten additions and the words in brackets are printed above the words between which they are placed here.

^{LXXV}Naumann, "Marcel Duchamp: A Reconciliation of Opposites," 30.

^{LXXVI}*Ibid.*

^{LXXVII}Quoted in Seigel, 164.

^{LXXVIII}*Ibid.*, 164-65.

^{LXXIX}Craig Adcock, "Duchamp's Eroticism: A Mathematical Analysis," in *Marcel Duchamp: Artist of the Century*, ed. Rudolf Kuenzli and Francis M. Naumann (Cambridge: MIT Press, 1987), 156.

The Effect of Islamic Azad University Activity on Branch of Trade Case Study: Firoozabad

Dr. Mehrnoush Saadat Abadi

Faculty of Islamic Azad University Firouzabad
147 St. John Street, 2nd Floor, London, UK
[Email: mseadat@yahoo.co.in](mailto:mseadat@yahoo.co.in)

Nima Naddaf (MBA) and Fateme Saeed Firouzabadi (BBA)

Faculty of Islamic Azad University Firouzabad
147 St. John Street, 2nd Floor, London, UK

Abstract

In this study we examine the impact of Islamic Azad University activities on Trade units of Firouzabad during 1991 to 2009 using multiple regression and Granger causality tests. Results show the expansion of university activities has positive effect on union significantly. Population growth and inflation have no effects on trade units at 95% confidence level. University activities expand are Granger causes of growth trade activities but population growth units are not. **Key words:** Islamic Azad University, Trade units, stationary, unit roots, Granger causality.

Considering the economical, social and cultural structure of province in Iran is major factor in disadvantaged provinces, mainly the lack of local skilled worker and infrastructure needs for recruitment expert. Therefore to achieve this goal and to make changes in the structure of society, planner needs to setup strengthening universities and higher education institutions in several development programs steps. As believed the major changes must base on staffing structure.

University is the institution that has special activities which brings value and credulity. The most important roles of universities are introducing research, education and social services. In other words, this agent is one of the most valuable resources for community development at its disposal. Today's economic system is based on competition and those who can continue to operate and remain in the competitive system have innovation activities. Obviously innovation requires creativity and without creativity access to the latest achievements of science - research is impossible. Therefore University plays a key role in developing And all of economic, social and cultural sectors rights to access to the latest research findings, should have close relationship with the university.

The purpose of this study is to analyze economic effect of Firouzabad Islamic Azad University on trade expansion unit.

The impact of University can be quantitative as well as qualitative. On the one hand increase the students, professor and staff led to a quantitative demand for goods and services. On the other hand increases the demand for high-quality of goods and services, expands production. Supplier units in both quality and quantity will expand its activities. Training provided by universities in the areas of management, business, economics and accounting have been the positive externalities on economic units, will forcing managers to optimize the supply of goods and services practices. One-way to expand the quality and quantity management activities of theses unit is the use of credit. Therefore in this research, banking credits that divided to these units use union and units as an indicator of trade expansion unit

Research Literature

Salehi, M. (2002) Research contributions in the force employed to separate primary, secondary and higher education in economic development with emphasis on the

force working with the college examined. In this study, using estimated production functions for calculate economic growth economy during 1966-96 has been discussed. . The results show that whatever labor is higher education, its effect on growth is stronger.

Komeyjany A. and Memarnejad, (2004) using a model of endogenous technological growth of the Romer (1990). They analyzed the positive effect of labor, human capital, physical capital, revenues from oil, the negative effect of inflation and changing Virtual Islamic Revolution to the rate of economic growth.

Results show that significant relationship was not view among the variable costs of research and development and growth.

Zara, M. and Ansari E. (2007) examine the relationship between and the cost of higher education and real GDP in. In this research three methods of standard Granger cause tests, Hyshao and (ARDL) method used to test this hypothesis. Review period is during 1974-2004. Results show that none of these variables cause another. Therefore, there was not relationship between economic growth and spending in higher education.

Ali Dehghani and others (2007) examine the effects of spending on research and development on the profit of cooperative companies and manufacturing industries of Razavi Khorasan province of Iran during 1995-2002. This information collected through 44 questionnaire spread in cooperative research and development unit of this province. Results suggest that research and development on all earnings-enhancing factors such as product innovation, low waste production problems and quality of goods is positive and significant.

Elmi, Z. and Jamshid Nejad, A (2007) study the effect of education on economic growth in the years 1972 to 2003 Iran. Theoretically, their analysis used Lucas model. In this study, the average years of education use as an indicator of force training and human capital.

Results show significant and positive effect of education on economic growth in Iran during the period reviewed.

Vaez, M. and others (2007) examines the effect of R & D costs and value-added industries with high technology during 1988-2006. In this study we use panel data collected for seven different industries. Result because of estimated show that research and development costs had important role in increasing the value-added of these industries.

Tayebi, K. and others (2008) studied the effect of human capital on economic growth of the Islamic Conference member countries during 2003-1980 using panel data. They apply econometrics model based on models followed

Sydrym and til. In this model independent variables which impact on economic growth are growth of industrial exports, growth imports of industrial, human capital development, workforce development and growth of physical capital. Model estimation results show that human capital growth has a significant and positive effect on economic growth.

Gyanaks Gerny. J. (2002) examined the role of human resources in economic growth using panel data over 98 countries. He showed that if when a dynamic model of panel data considered, training has a positive and significant effect on economic growth. In addition, the effect of higher levels of education on economic growth is stronger.

Kaiser, A. (2002) used data for the years 1970-1994 to study the impact of education on growth. Investigation results show the primary school as an indicator of human capital has negative effect on growth for the Pakistan and Sri Lanka. And if in secondary school education is, its effect on economic growth for the two countries will be positive and significant. Education above the high school similarly has a significant and positive effect on economic growth and development.

Theoretical Foundations

The purpose of this study is review the relationship between bank credit growth and the number of students enrolled each year. Theoretical Foundations of statistical methods of research and analysis tools present in this section. To delete the effect of inflation on credit, we use the actual of funds that earned by dividing nominal credit on consumption expenses price index (CPI).

The credit data collected from the Firouzabad banking.

Index of population growth calculated by difference of born and died during these years. The data got from the Bureau registration of Firouzabad.

Unit root test:

Many macroeconomic variables in time series are non-stationary (Hill and others: 2001). If a series is stationary, then the shock imposed on the Mira variables Destruct and returns to the long-term. On the other hand, if the time series is non-stationary, the mean or variance or both of them are a function of time. And if the time is infinite, varying the variables will be infinite. Therefore, the variable will divergent away from its path equilibrium (Stereo and Hall: 2006). Philippe - Peron test (1988) used for study of stationary test. In this test, the existence of unit root for time series in question is hypothesis. In case of stationary variables, using ordinary least squares estimation is desirable.

Granger causality test

If the lag values of x, effect on the current values of y, in other words changes in y looking to create change in x, then x called Granger causes of y.

Regression presented in

$$y_t = \beta_0 + \sum_{i=1}^k \beta_i y_{t-i} + \sum_{i=1}^k \gamma_i x_{t-1} + \epsilon_t$$

Zero hypothesis is performed on F-test as follows.

$$H_0 : \gamma_i = 0$$

If the zero hypotheses are rejected, then x is Granger causes of y.

If variables are integrated of order one, first order Difference Will transform them to non- integrated. Then above test apply in the following form.

$$\Delta x_t = a_x + \sum_{i=1}^k \beta_{x,i} \Delta x_{t-i} + \sum_{i=1}^k \gamma_{x,i} \Delta y_{t-i} +$$

$$\Delta y_t = a_y + \sum_{i=1}^k \beta_{y,i} \Delta y_{t-i} + \sum_{i=1}^k \gamma_{y,i} \Delta x_{t-i} +$$

Δx_t and Δy_t are the first order differencing of time series that non- integration

Research Findings

In this section, using statistical tools and econometric techniques, the interaction of variables will be analyzed and the results will be presented.

Period of study is 1371-1388. For data analysis we use Eviews.6 software.

First describe the statistical variables in question. These results are in Table -1.

Table 1- statistical describes of variables

	Number of registered students (student)	Population (pop)	Bank credit (credit)
Mean	667.2138	141771.4	667.4315
Median	000.2720	990610.3	000.4512
Maximum	000.5129	343413.7	000.6544
Minimum	000.3884-	341352.0	0000.888
Standard deviation	697.1771	623194.1	697.1771
Sum	00.32080	12657.62	00.64735
Sum of squares of deviations	91161357	88663.36	43944723
Observation	15	15	15

Source: research findings

As stated in the previous section, Philip - Peron test, (1988) use for the study of stationary series. The hypothesis tests are the existence of unit root for time series variables which studied.

Unit root test results for variables in levels show in Table -2.

Table -2: Unit root test results

Variables	Level		
	Probability	t-statistic	Critical value at 5%
credit	0476.0	759743.3-	759743/3-
student	0.0156	-3.660851	-3.052169
pop	0763.0	749065.1-	-1.962813

Source: research findings

Test results show that the evidence unit root in these variables are not observed. Therefore, variables are stationary and put to analysis in the Multiple regression method In this section, the nominal credit as the dependent variable, the number of students enrolled and the consumer price index as independent variables use in the regression equations. Estimated result present in Table -3.

Table -3 estimated parameters

Independent Variable Credit				
dependent variable	Coefficients	t-statistic	Standard error	Probability
Constant	-93.06908	-.513060	181.4001	0.6172
student	0.198328	2.243714	0.088393	0.0445

Independent Variable Credit				
dependent variable	Coefficients	t-statistic	Standard error	Probability
CPI	-1.567821	-1.018118	1.539921	0.3287
R ² =0.571997				

Source: research findings

The results show the coefficient of (student) is positive and significant at 95% confidence level.

Therefore we can end that increasing university activities, may lead to increase in bank credit.

On the other hand the coefficient of consumer price index variable index (CPI) in the 95% confidence level is not significant. So would not suggest that increase in inflation lead to increase in funds.

Then, we use changes in bank credit as the dependent variable and changes in student and population as dependent variable. Results are present in table 4.

Coefficient of changes in students (Dstudent) is positive and significant at 95% confidence. So, again evidence shows a positive relationship between these two variables. On the other hand, coefficient of (Dpop) is not significant at 95% confidence level. Therefore, can not say this variable affects credit significantly.

Table -4 estimate parameters

Independent Variable Credit				
dependent variable	Coefficients	t-statistic	Standard error	Probability
Constant	-1.693836	-2.932336	0.577640	0.0136
D student	0.003822	3.011245	0.001269	0.0118
D pop	0.000243	1.717479	0.000141	0.1139
R ² =0.453676				

Source: research findings

Granger Causality Test

To check the relationship between bank credit union assigned as an indicator of growth units and number of students registered as growth University activities indicator, we use Granger causality test.

Granger- causality test results between variables are in Table -5.

Results reject the hypothesis of students enrolled is not Granger- causality of bank credit.

Therefore we can say increasing the number of students, increased bank credit allocated to the unit's trade in FirouzAbad which consistent whit results of regression method of above.

Hypothesis of population growth is not Granger-causality of credit does not reject, so the only variables affect students are enrolled. The results show as student increase, the population city increase.

Table -5 Granger causality test results

H ₀ Hypothesis	Observation	F-statistic	Probability
students registered is not Granger-causality of bank credit	13	4.30323	0.0539
Bank credit is not Granger-causality of students registered.	13	1.39145	0.3030
Population growth is not Granger- causality of Bank credit	15	2.84001	0.1055
Bank credit is not Granger- causality of Population growth	15	2.19668	0.1619
Population growth is not Granger- causality of student registered	15	2.39164	0.1416
student registered is not Granger- causality of Population growth	15	7.56724	0.0100

Source: research findings

Conclusion

Nowadays increasing labor productivity and economic development planning are agenda in developing countries.

One of the important ways to increase workforce productivity and quality of labor force is through education and creating favorable grounds and easy access to education for various social groups in various fields. Basic education development in the individual governing values, attitude and behavior, and he creates the groundwork for acceptance of new management practices and production business provides. One of the basic tools in university extension education development regions in particular is less developed regions. Islamic Azad University, a successful example of this process is to grow and expand economic activities in different regions has contributed.

The impact of University can be quantitative as well as qualitative. On the one hand increase the students,

professor and staff led to a quantitative demand for goods and services. On the other hand increases the demand for high-quality of goods and services, expands production. Supplier units in both quality and quantity will expand its activities. Training provided by universities in the areas of management, business, economics and accounting have been the positive externalities on economic units, will forcing managers to optimize the supply of goods and services practices.

In this study we analysis the effects of Firouzabad Islamic Azad University on economic growth. Data which use are during 1992 to 2009. Multiple regression techniques and Granger causality test are applied. The relationship between growth of students enrolled as an indicator of University activities and bank credit allocated to the unit as an indicator of regional trade are examined. Effect of population growth and inflation rates on credit growth analyzed.

Regression results indicate that the effect of university activities on the corporate unit growth is positive and significant at the level of 95% confidence.

But population growth and inflation significant effect on union activity is units. Granger-causality results confirming the above regression model.

References

- Andrew Levin, Lakshmi K. Raut (1992) "Complementarities between Exports and Human Capital in Economic Growth: Evidence from the Semi-industrialized Countries, Economic Development and Cultural Change
- Antonio Ciccone, Elias Papaioannou (2006) "Human Capital, The Structure of Production Research and Growth" Working Paper Series, NO 623 / May 2006
- Dehghani, Ali (2007) "Cost effectiveness of R & D (case study of production cooperatives and industries of Iran Khorasan Razavi province)," Economic Quarterly Reviews Volume 4 Issue 2 Summer 2007 2
- Elmi, Z. Jamshid Nejad, A (2007) " Effect of education on economic growth in Iran in 1971-2003 " Journal of Human Sciences Special Economic seventh year, No. 26, Fall 2007
- G. Agiomirginakis, D. Asteroid, and V, Monastrriotis. (2002) "Human Capital and Economic Growth Revisited: A Dynamic Panel Data Study" IAER: August 2002, VOL. 8, NO. 3
- Komyjany, A. and Memar nejad A. (2004) "the importance of quality manpower and R & D (research and development) on Iran's economic growth" Quarterly Journal of Business, No. 31, Summer 2004
- Michael Boozer, Gustav Rains, Frances Stewart (2003) "Paths to Success: The Relationship Between Human Development and Economic Growth" Economic Growth Center, Yale University, Center Discussion Paper NO. 874
- Qaisar Abbas (2001) "Endogenous Growth and Human Capital: A Comparative Study of Pakistan and Sri Lanka "The Pakistan Development Review40: 4 Part II (Winter 2001) pp. 987-100
- Raul, R. Surinach, J. Artís, M. (2009) "Regional Economic Growth and Human Capital: The Role of Over education" Discussion Paper Series, IZA DP No. 4453
- Salehi, M.(2003) "Human Capital Effects on Economic Growth" Journal of Research and Planning in Higher Education Letter, No. 22 and 24
- Tayebi, K. and others (2008) "Effect of Foreign Trade and human capital on economic growth in OIC countries," the source www.sid.ir
- Vaez, M. and others (2007) "The role of research and development costs in high value industries added technology" Quarterly Economic Review, Volume 4, Number 4, Winter 2007
- Zra', M. and Ansari, E. (2007) "Granger causality of the relationship between economic growth and the cost of higher education in Iran," Iranian Journal of Economic Research, No. 37, Winter 2008 Twelfth Year

Service Failure and Recovery: Comparison Between Health Care and Automobile Service Station

Anupam Krishna, G.S. Dangayach, Rakesh Jain

Department of Management Studies
Malaviya National Institute of Technology, Jaipur, India

Abstract

Purpose – The paper aims to give insights into customers' perceptions and response regarding service failure and recovery process for health care and automobile service station. It compares and recommends the service recovery strategies between the two to gain higher satisfaction, trust and loyalty. **Design/methodology/approach** – The paper is empirical and involves data from 150 respondents from India. The sampling unit is customers of health care and automobile service station and the data collection instrument is a structured, non-disguised questionnaire. The questionnaires have been filled through individual interviews. **Findings** – The study found conclusive results on the reasons for service failure and recovery strategies in the health care and automobile service station business in the Indian context. It also lays emphasis on the behavioral aspects of customers' perception towards these failures and their responses to the same. The paper further looked into the recovery strategy employed by health care and automobile service station and the customers' perception towards the recovery strategies. **Practical implications** – The study has important managerial implications as it facilitates the understanding of failure – points and its occurrence. The comparative study shows, how the two sectors differ for similar category of failure in context to, severity of failure and recovery strategy for it. The paper is also helpful in designing policies and procedures to proactively eliminate such failure points. **Originality/value** – The paper is valuable as very little has been done in the Indian context. It also creates a comparative perspective as to service failure in the health care and automobile service station sector. **Key words:** Health Care and Automobile Service Station, Customer service management, Customer satisfaction, Consumer behavior.

With the onset of globalization and liberalization, the business competition is growing exponentially, which caters to strong need of improving the overall quality (Zeithaml et al., 1988, Lee et al., 2003). Increase in service sectors contribution to the business has been seen globally. The need of hour is to produce high quality (Swan and Bowers 1998, Michael and McCollough 2000) and defect free services (Lovelock and Wirtz, 2004). Various researches have been done on service sector to find the buyer seller gaps, customer perceptions and expectation, because fulfillment of expectation confirms satisfaction (Oliver, 1980, Parasuraman et al., 1988; Zeithaml, 1988; Sureshchandar et al., 2003; O'Neill and Palmer, 2004). The studies on outcome and process service failure had contributed to find the cause and effect of failure. Business people have been always looking for some technique to proactively know about the customer perception and expectancy to eliminate the failure. But so far no technique could be designed to provide all together error free services. The academicians and practitioners know that the despite all efforts, error may still creep in (Mack et al., 2000). Now in his situation the tool called service recovery becomes very important.

Service recovery is a process to eliminate or reduce the effects of failures by compensating or supporting for the loss of customer. The service recovery not only a tool to reduce harmful effects of failure but it can also be useful to gain satisfaction, trust and loyalty. Moreover researchers have identified it as tool for endorsement to enhance customer base (Ruyter et al 1998).

Service Recovery

Service recovery is referred to as the action undertaken by an organization to face the eventualities of a service failure (Zeithaml and Bitner, 2000). Boshoff, (1997) suggests that, a fast response by the highest possible person in terms of seniority with a fast response accompanied by a full/partial refund plus some amount of compensation can reduce the effects of failures. Gronroos 1988 The researchers have produced three kinds of justice that can be used after the failure, which are, fairness of the resolution procedures (procedural justice), the interpersonal communications and behaviors (interactional justice), and the outcomes (distributive justice). Informational justice

is newly added to the recovery choices which consists of complaint handling process including elements such as politeness and courtesy exhibited by personnel, empathy, effort observed in resolving the situation, and the firm's willingness to provide an explanation as to find out, why the failure occurred. Bell and Zemke (1987) proposed five dimensions for recovery i.e. Apology, Empathy, Urgent reinstatement, Symbolic atonement and Follow-up, Bitner et al. (1990) gave the process of recovery in four steps.

- Problem Acknowledgement
- Explanation of the reason
- Apology where appropriate
- Compensation such a free ticket, discount coupons etc.

The process of recovery is as below:

1. Acknowledgement: Acknowledging that a problem has occurred (Bitner et al, 1990).
2. Empathy: Understanding the problem from a customer's point of view (Johnston and Fern, 1999).
3. Apology: Just Saying sorry (Kelley et al., 1993).
4. Own the problem: Taking ownership of the customer and the issue (Barlow and Møller, 1996).
5. Fix the problem: Fixing, or at least trying to fix the problem for the customer (Michel, 2004).
6. Provide assurance: Providing assurance that the problem has been/will be sorted and should not occur again (Barlow and Møller, 1996).
7. Provide compensation: Providing a refund, and/or a token and/or compensation, depending on the severity of the problem (Boshoff, 1997).

Service recovery quality main ingredients are assurance, reliability, facilities, employee's empowerment, customization, and responsiveness (Gilbert and Wong 2003). Service recovery also infuses loyalty and trust by improving long-term customer retention (Mueller et al.). Timely recovery with the correction of problem is necessary for successful service recovery (Wirtz & Mattila, 2004). Researches have shown that, customers may accept failures, but may not forgive organizations that can't or won't fix them' (Mattila, 2006).

The Current Study

The objective of the study is to conduct a comparative analysis of service-failure causes and recovery strategies in health care and automobile service station in the Indian context. The objective of the study is to compare the consumer experience while taking services from health

care sector such as hospitals, clinics etc and automobile service station (four and two wheeler only) in India. The study intends to determine if there is major difference in consumer behavior of these two kinds of service sectors to draw significant conclusions and indicators for hospitality managers in these kinds of businesses.

The following objectives have been defined for this study:

- To identify causes of usual service failure in health care and automobile service station.
- To understand specific complaints and categorize them into generic classes
- To identify usual strategies health care and automobile service station adopt to cope with such failures and
- To measure the outcome of these service recovery strategies and evaluate the impact of the chosen strategy on consumer perception and future behavior.

This study entailed data collection in India from health care sector such as hospitals and clinics and automobile service station. The data collection instrument used is a structured and non-disguised questionnaire. The questionnaire was formulated on the basis of suggestion by subject matter specialists in India.

The different kind of service-failures and corresponding recovery strategies that have been included in the questionnaire for this study are based on review of extant literature. Berry et al. (1990) finds that whenever a customer experiences a problem his confidence, trust and risk taking with provider reduces. The response of provider after the failure decides that the customer could be physically and as well as mentally retained or not. It is possible that unsatisfied customer may be still giving the business but due to grudge holding, whenever he gets chance he may switch off.

The current study has tried to address following issues and present appropriate findings.

1. In the case of similar kind of failure in different service industry, does the severity of failure according to the customer perception differ?
2. In the case of similar kind of failure in different service industry, does the expectancy of service recovery differ according to the customer perception?
3. The customer may not be willing to return to provider even after the expected recovery, In the case of similar kind of failure in different service industry

Researcher have done a variety of experimental design, field studies and empirical studies that examine the effects

of service recovery on satisfaction, purchase intentions (Ruyter et al 1998) and one's tendency to spread positive word-of-mouth information. To get the expected recovery prediction on consumer response is must (Maxham, 2001; and Flokes, 1984). Likely behavior depends on Customer's prior experiences and expectations which may be Exit, switch, Voice, negative word of mouth publicity (Kau and Wan-Yiun Loh 2006, Singh 1990). Zeithaml et al produced model which specifies three different types of service expectations: desired service, adequate service, and predicted service. Researchers keep on trying to find out ways to avoid switching (Colgate and Lang 2001). Counterfactual thinking (Walchi and Landman 2003), anger (Bougie et al. 2003), customer complaints, reciprocity and complaint handling and ways to instill a service recovery strategy are of vital important in service recovery process. For effective recovery process complains management has to be effective because, complaining customers are among the most loyal customers (Eccles and Durand 1998). Services and service recovery should be designed according to human behavioral science principles underlying human interactions that can be translated directly into service design. It has been found that behavior changes with the age, sex, culture, occupation, and experiences (Patterson 2007). Jones and Farquhar (2007) studied the behavioral intension in different failure magnitudes.

Weun et al. (2004) have found that if the original problem was severe there remained a negative influence on customer's satisfaction, even if there is effective service recovery. In this study we have tried to find that with same kind of failure in two different sectors, the customer is willing to repeat his purchase in the future or not.

The questionnaires were administered personally to a sample of respondents in both the sectors. Convenience sampling was adopted to collect responses from one hundred and fifty respondents who faced service failures, which was considered to be a large enough sample for an exploratory study of this nature (Malhotra, 2001). To complete this target of 150 respondents individuals were contacted as only those who have experienced service-failure of any kind in a health care and automobile service station were the sample population. Data analysis entailed forming comparative statements of both markets to determine trends. Based on literature review the major types of service failure in health care and automobile service station have been identified in three broad categories.

Hygiene and Physical Evidence- poor cleanliness, untidy staff, not appealing (looks), poor ambience, stingy,

water and facilities arrangement not proper, congestion, arrangement in waiting room not proper, uncomfortable temperature.

Operations- slow service, specialist unavailability, ambiguity of rate, ambiguity in process/ waiting time, machine or equipment unavailability, wrongly charged, work not done proper, lost order, missing of your personal items, reservation missing.

Employee related – untidy staff, staff not prompt, staff not attentive, lacking in efforts, does not understand your needs, staff does not have knowledge about their jobs, unfriendly and unhelpful staff

Likewise, different strategies used for failure recovery by health care and automobile service station have been identified and incorporated in the questionnaire. The effects of various recovery strategies have also been evaluated in the form of customer response in the form of repeat visits and recommendations to people and associates, and the same has been used to sketch managerial implications. The study has an urban prejudice as it was conducted in cities like Delhi, Jaipur, and Lucknow.

Findings and Analysis

Customer Complaint Behavior

The data indicates that a wide range of failures in both the sectors exists due to large number of reasons. The two sectors differ in number of people who reported failures. It was found that in automobile service 97% of people compared to health care only 56% consumers have experienced service failure of one or another kind. The difference between the two sector is that in automobile service station almost everyone has made a verbal compliant, only 61% of the health care chose to register their complain. This may be explained as people may find it is not the place to raise their voice and should maintain its dignity and decorum. This research finds that implementation of complain management in India is under infancy stage. Generally we don't find any service recovery team, service recovery policy for the failure. Indian customer believes that the magnitude of verbalization of complain decides the promptness and degree of recovery. Consumer believes that unless you verbally complain, no action would be taken by the provider. It has been found in this study that customer do not have much trust on provider which is a serious issue.

Categories of Service Failure

The study has found generic resemblance across the two service sector segments as far as causes for dissatisfactions are concerned. This study finds that, most of the service failures were common in both the sectors and can be categorized in three main headings those are:

1. Hygiene and Physical Evidence- poor cleanliness, untidy staff, not appealing (looks), poor ambience, stingy, water and facilities arrangement not proper, congestion, arrangement in waiting room not proper, uncomfortable temperature.
2. Operations- slow service, specialist unavailability, ambiguity of rate, ambiguity in process/ waiting time, machine or equipment unavailability, wrongly charged, work not done proper, lost order, missing of your personal items, reservation missing.
3. Employee related – untidy staff, staff not prompt, staff not attentive, lacking in efforts, does not understand your needs, staff does not have knowledge about their jobs, unfriendly and unhelpful staff

The operations related failures were considered the most significant in relation to customer complain behavior. Then in the rank comes employee related, followed by hygiene and physical evidence category. The rank according to the commonness of the specific failure is shown in

Table 1

Failures	Severity Health Care (%)	Rank	Severity Automobile service (%)	Rank
Hygiene and Physical Evidence				
Poor cleanliness	15.1	13	3.4	19
Untidy staff	10.2	14	00	-
Not appealing (looks)	6.6	16	00	-
Poor ambience	16.3	12	00	-
Stingy	22.6	7	00	-
Water and facilities arrangement not proper	4.5	19	9.4	16
Congestion	4.6	18	2.5	20
Arrangement in waiting room not proper	00	-	10.5	15
Uncomfortable temperature	3.2	20	4.8	18
Operations				
Slow service	67.7	1	70.5	1
Specialist unavailability	61.2	2	68.6	2
Ambiguity of prices	00	-	15.7	12
Ambiguity in process/ waiting time	00	-	21.7	10
Machine or equipment unavailability	34.8	4	37.7	7

Failures	Severity Health Care (%)	Rank	Severity Automobile service (%)	Rank
Wrongly charged	4.7	17	40.1	5
Work not done proper	22.4	8	44.7	4
Lost order	00	-	12.7	14
Missing of your personal items	00	-	13.8	12
Reservation missing	00	-	8.4	17
Employee related				
Untidy staff	6.8	15	00	-
Staff not prompt	34.5	5	23.8	9
Staff not attentive	21.5	9	17.5	12
Lacking in efforts	36.6	3	38.8	6
Does not understand your needs	26.6	7	35.6	8
Staff doesn't have knowledge of their jobs	31.7	6	50.5	3
Unfriendly and unhelpful staff	18.4	10	13.8	13

Table 1: Customer perceptions about intensity of service failure

It was found that for the same category of failure the intensity was different. In health care segment customer percieveance for the failure seems to be serious as compared with automobile service segment. Reason for this can be explained as in the health care, the health of an individual is in question, which is invaluable and cannot be compared to the value of an automobile. About 82 percent respondents feel that the intensity of the problem they encountered in the health care sector is extremely serious whereas for automobile service station customers this has been around 31 percent. Health care segment consumer's expectancy of error free service is more as compared to the other sector. Even for the low intensity error in service station such as cleanliness and inattentive employee they take this as high intensity failure. It is also noted that health care sector pays more attention for delivering error free services.

In cities like New Delhi, Jaipur, and Lucknow, we find mushrooming populations; escalating cost of living overcrowded spaces, long-drawn-out public services and growing middle class population. This has created unhealthy competition, insecurity, tension and stress in individual and relation between common people. Its reflection can be seen in the buyer seller relation and service process too. Many times customer yells at the frontline and managers to vent their frustration and grudge from other situations. It becomes difficult for frontline employees to manage such situation. From the management they have pressure to be defensive, this creates job dissatisfaction and internal frustration, which comes back to some other customer in some other form.

Service Failure Recovery Expectancy Vs. Action Planned

The research finds that adequate service recovery process and policy is lacking. It is observed that front line employees have a general tendency to transfer the case to higher authority. Possible reason for this may be lack of employee empowerment. It is also observed that in both sectors generally providers perceive recovery process as a time and money waste process. Even the middle and top level people were found to misunderstand the service recovery process. Despite of finding real reason, generally complain reason is straight away taken as inefficient front line employee, so we see discouragement of complaint behavior from the provider. Almost 40 percent of complaints in the Indian context are dealt by extending an apology and offering short compensation. In health care sector, we find 75 percent cases of similar type.

The expected recovery and received recovery were found to be quite different; in case of health care the general expectancy was sympathy, empathy, and apology and owning the problem. Study finds that provider's way of recovery was to give apology and short compensation. It was observed that offering compensation for health loss sometimes may be offensive to the consumer, as we cannot evaluate someone's life or health in terms of money. In the case of automobile service station expected recovery on the first choice was compensation in cash or kind then other choices were explanation and empathy. In this case generally the received recovery was explanation and apology. This sector was found to resist for cash or kind compensation, as very few cases reported for compensation in cash or kind and that too very short. It has been observed that cash or kind compensation works better to any other strategy in automobile service segment and apology and sympathy works better in the case of health care segment.

The most important aspect of this study was the consumer behavior after the service recovery process. In automobile service station prompt action after failure was reported 65 %, whereas in it health care segment was reported 54 %. It was also noted that about 71% people of automobile segment believed that recovery could be better than the received one whereas only 46% people believed it in the health care segment. A total of 61% of health sector and 23% of service station sector reported having bad memories of failure and unsatisfied recovery action. Customers who were willing to give business to provider were 22 % in health care and 68% in case of automobile service segment. To suggest provider to someone 13% was found in health care and 24%

in the case of automobile service segment. This indicates loss of clientele and revenue of provider. Only 6% of health care sector and 28% of automobile service segment reported to have positive memory after the recovery. It was observed that customer of automobile station believed error may occur and service failure may happen but in case of health care segment consumer do not expect any sort of service failure. This is due the basic fact that consumer of health care know that just a small error can question some bodies life whereas this in the service station chances of serious loss are low. This indicates that occurrence service recovery paradox has greater chances in the sectors like service station where the loss can be compensated.

Managerial Implications

This study points out a number of managerial implications. This research clearly indicates that the service intensity and severity has lot to do with the type of sector which it is associated with. The consumer perception about the failure heavily depends on the kind of loss he is going to suffer in case of failure. Study identifies that whatever kind of failure it is, basically it comprises of certain kind of losses i.e. Health loss, Character loss, Emotional loss, Respect loss, Loss of identity/image, Relationship loss, Loss of belief, Time loss, Money loss, Comfort loss, Mental/physical energy loss, and Loss of control. Study observes that health loss carries more importance than money or time loss. Hence the avenues related to or their failure outcome could be health loss should be more attentive to the failure as customers do not expect failures from them. This study further finds that the recovery in case of healthcare loss is difficult to recover than service station failure. Recovering health or respect loss kind of failure with cash compensation can be disastrous. Hence this paper suggests that service recovery paradox occurrence in health care segment is difficult as providing effective recovery is not easy.

This study observes that Indian consumer lack in feeling of trust for the providers. They believe that the provider will not take action until you yell the problem to them, and the provider recovery action intensity is proportional to intensity of yelling. This is a serious issue and Trust could be gained by making long term policies in complain management and recovery process. This study recommends training and motivational programs for employees to learn what and how is react in case of failures. Further Indian providers should not blame frontline employees for every kind of error but they should empower them to take appropriate action in failure cases rather to shift the care to senior people. The

satisfied employee would not only show his efficiency in delivery but also in service recovery and customer building. This study observes that a concrete complain management and service recovery policy is lacking in Indian context. Service provider need to understand that they are not time and money waste instruments, but are vital tools for building business. Service recovery not only retains clients but builds new clients too by positive word of mouth effect from satisfied customers.

References

- Andreasen, A.R. and Best, A. (1977), "Consumers complain— does business respond?", *Harvard Business Review*, July/August, pp. 93-101.
- Andreassen T W (1998), "Antecedents to satisfaction with service recovery", *European Journal of Marketing*, Vol. 34 No. 1/2 156-175 pp.
- Andreassen, T.W, (1999), "What drives customer loyalty with complaint resolution?", *Journal of Service Research*, Vol. 1 No. 4, pp. 324-32.
- Anna S. Mattila (1999), "An Examination of Factors Affecting Service Recovery in a Restaurant Setting", *Journal of Hospitality & Tourism Research*, vol. 23: pp. 284 – 298.
- Bailey, D. (1994), "Recovery from customer service shortfalls", *Managing Service Quality*, Vol. 4 No. 6, pp. 25-8.
- Bernd Stauss (2002), "The dimensions of complaint satisfaction: process and outcome complaint satisfaction versus cold fact and warm act complaint satisfaction", *Managing Service Quality*, Volume: 12 Issue: 3 Page: 173 – 183.
- Berry, L.L, V A Zeithaml and A Parasuraman (1990), "Five Imperatives for Improving Service Quality", *Sloan Management Review*, vol. 31, no 4 Summer (1990), pp 29-38.
- Bies, R. and Shapiro, D. (1987), "Interactional fairness judgments: the influence of causal accounts", *Social Justice Research*, Vol. 1 No. 2, pp. (199-218).
- Bell, C.R. and Ridge, K. (1992) "Service recovery for trainers", *Training & Development*, Vol. 46, p. 58.
- Bell, C.R. and Zemke, R.E. (1987), "Service breakdown: the road to recovery", *Management Review*, October, pp. 32-5.
- Blodgett, J.G., Hill, D.J. and Tax, S.S. (1997), "The effects of distributive, procedural and interactional justice on post-complaint behavior", *Journal of Retailing*, Vol. 73 No. 2, pp. 185-210.
- Chandrashekar, Murali, Kristin Rotte, Stephen S. Tax, and Rajdeep Grewal (2007), "Satisfaction Strength and Customer Loyalty", *Journal of Marketing Research*, 44 (1), 153-63.
- Christo Boshoff (1997), "An experimental study of service recovery options", *International Journal of Service Industry Management*, Volume: 8 Issue: 2 Page: 110 – 130.
- Christo Boshoff (2005), "A re-assessment and refinement of RECOVSAT An instrument to measure satisfaction with transaction-specific service recovery", *Managing Service Quality*, Vol. 15 No. 5, pp. 410-425.
- Colgate, M. and Lang, B (2001), "Switching barriers in consumer markets: an investigation of the financial services industry", *Journal of Consumer Marketing*, Vol. 18 No. 4, pp. 332-47.
- Craighead, C.W., Karwan, K.R. and Miller, J.L.(2004), "The effects of severity of failure and customer loyalty on service recovery strategies", *Production and Operations Management*, Vol. 13 No. 4, pp. 307-21.
- David Cranage (2004), "Plan to do it right: and plan for recovery", *International Journal of Contemporary Hospitality Management*, Volume 16 • Number 4 • pp. 210-219.
- Davidow, M. (2003), "Organizational responses to customer complaints: what works and what doesn't", *Journal of Service Research*, Vol. 5 No. 3, pp. 225-50.
- Etzel, Michael J. and Bernard I. Silverman (1981), "A Managerial Perspective on Directions for Retail Customer Dissatisfaction Research", *Journal of Retailing*, 57 (3), 124-36.
- Feinberg, R.A., Widdows, R., Hirsch-Wyncott, M., Trappey, C. (1990), "Myth and reality in customer service: good and bad service sometimes leads to repurchase", *Journal of Consumer Satisfaction Dissatisfaction and Complaining Behavior*, Vol. 3 pp.112-4.
- Folkes, V.S. and Kotsos, B.(1986), "Buyers' and sellers' explanations for product failure: who done it?", *Journal of Marketing*, Vol. 50, April, pp. 74-80.
- Gavin Eccles G. and Durand P. (1998), "Complaining customers, service recovery and continuous improvement", *Managing Service Quality*, Volume 8 • Number 1 • pp. 68–71.

- Gilly, M. C. (1987), "Post complaint processes: From organization response to repurchase behavior.", *Journal of Consumer Affairs*, 21(2), 293-313.
- Gremler, D.D (2004), "The critical incident technique in service research", *Journal of Service Research*, Vol. 7 No. 1, pp. 65-89.
- Groenroos, C. (1984), "A service quality model and its marketing implications", *European Journal of Marketing*, Vol. 18 No. 4, pp. 36-44.
- Gronroos (1990), "relation approach to marketing in service context the marketing and organisation interphase", *journal of business research*.
- Gronroos, C. (1988), "Service quality: The six criteria of good perceived service quality, *Review of business*, Vol. 9 No.3, pp.10-13.
- Halstead, D. and Page, T.J. (1992) "The effects of satisfaction and complaining behavior on consumer repurchase intentions", *Journal of Customer Satisfaction Dissatisfaction and Complaining Behavior*, Vol. 5, pp. 1-11.
- Hart, C.W.L., Heskett, J.L. and Sasser, W.E. (1990), "The profitable art of service recovery", *Harvard Business Review*, Vol. 68 No. 4, July-August, pp. 148-56.
- Kelley, S.W. and Davis, M.A. (1994), "Antecedents to customer expectations to service recovery", *Journal of the Academy of Marketing Science*, Vol. 22 Winter, pp. 52-61.
- Lapidus, R.S. and Pinkerton, L. (1995), "Customer complaint situations: an equity theory perspective", *Psychology & Marketing*, Vol. 12 No. 2, pp. 105-22.
- Maxham, J.G.I. and Netemeyer, R.G. (2003), "Firms reap what they sow: the effects of shared values and perceived organizational justice on customers' evaluation of complaint handling", *Journal of Marketing*, Vol. 67 No. 1, pp. 29-45.
- McCollough, M.A. and Bharadwaj, S.G. " ", in Allen, C.T. (Ed.). (1992), "The recovery paradox: an examination of customer satisfaction in relation to disconfirmation, service quality, and attribution based theories", *Marketing Theory and Applications*, American Marketing Association, Chicago., IL, p. 1(19. Michel S (2001), "Analyzing service failures and recoveries a process approach", *International Journal of Service Industry Management*, 12,1 pp. 20-33.
- Oliver, R. (1993), "Cognitive, affective, and attribute bases of the satisfaction response", *Journal of Consumer Research*, Vol. 20 No. 1, pp. 418-30.
- Parasuraman, A., Zeithaml, V. and Berry, L.L. (1985), "A conceptual model of service quality and its implications for future research", *Journal of Marketing*, Vol. 49, Fall, pp. 41-50.
- Reichheld, Frederick F.(1994), "Loyalty and the Renaissance of Marketing", *Marketing Management*, 2 (4), 10-21.
- Singh, J. (1990), "Voice, exit and negative word-of-mouth behaviors: an investigation across three service categories", *Journal of the Academy of Marketing Science*, Vol. 18 No. 1, pp. 1-15.
- Smith, A.K. and Bolton, R.N. (1998), "An experimental investigation of customer reactions to service failures: paradox or peril?", *Journal of Service Research*, Vol. 1 No. 1, pp. 65-81.
- Tax, S.S. and Brown, S.W. (1998), "Recovering and learning from service failures", *Sloan Management Review*, Fall, pp. 75-88.
- Zeithaml, V.A., Berry, L.L. and Parasuraman, A. (1996), "The behavioral consequences of service quality", *Journal of Marketing*, Vol. 60, pp. 31-46
- Zhu, Z., Sivakumar, K. and Parasuraman, A. (2004), "A mathematical model of service failure and recovery strategies", *Decision Sciences*, Vol. 35