Development of a Data Visualization Model based on Information Processing Theory

Falschlunger Lisa[1], Lehner Othmar, Eisl Christoph and Losbichler Heimo

University of Applied Sciences Upper Austria, Faculty for Management Steyr
Wehrgrabengasse 1-3, 4400 Steyr, AUTRIA

1 Corresponding author's email: lisa.falschlunger@fh-steyr.at

ABSTRACT:
Information has emerged as being the fourth production factor and is becoming increasingly important in global competition. Research shows that managerial decision making is directly correlated to both, the swift availability, and subsequently the ease of interpretation of the relevant information. A holistic picture, including the relevant key performance indicators of the company, zooming between the micro and the macro environments, as well as the ability of the involved managers to interpret this information is therefore crucial for the performance of a company. The availability of so called 'big data' and the ever shorter cycle time between the identification of an important piece of information and the possible need of the management to react force traditional business concepts to change. Visualizations are already widely used to transform raw data into a more understandable format and to compress the huge amounts of data produced. However, research in this area is highly fragmented and results are contradicting. This paper proposes a preliminary model based on an extensive literature review incl. top current research on cognition theory. Furthermore an early stage validation of this model by experimental research using structural equation modeling is presented. The authors are able to identify predicting and moderating variables for information perception of visual data.

1 INTRODUCTION

A visual representation of data can be seen as a means to accelerate, as well as to improve cognition and interpretation [1] of such, and thus should in theory improve rational managerial decision making. However, such representations so far are used inconsistently in praxis [4] [6] and sometimes either in a way that misses the purpose of informing the reader in an effective and efficient way [9] or in a way that may be even misleading or manipulating [3].

According to Conati and Maclaren [5] the success of visualizations should be determined by the ability of users to retrieve relevant information in an effective and efficient way. In this context the theory of cognitive fit has been used as a research framework, however, results are somewhat contradicting, often due to the lack of a holistic knowledge in relation to the visual perception process (e.g.: [2] [7] [10] [22] [26]).

How visualizations affect a user's perception is a highly complex process as it is influenced by the task and the data at hand as well as by individual factors such as experience, knowledge or culture [20] [6] [21] [32]. So far, research in the area of information visualization lacks a means and method of determining the quality of a graphical representation for such purposes *a-priori* [33] [21] taking contextual factors into account.

To advance theory, the authors introduce a triangulated and validated holistic model based on structural equation modeling to allow the a-priori evaluation of cognitive effectivity and efficiency, while remaining adaptive to individual recipients' differences by incorporating the cognitive burden a visualization presents to its reader.

This paper is structured as follows: In the next section there is an outline of the theoretical foundations of information visualization before stating the research questions in the third section. Part four introduces the methodology and the last section outlines preliminary results of the developed model.

2 THEORETICAL BACKGROUND

The model is based on information processing theory and its further stage cognitive load theory. Information processing theory provides knowledge on the perceptual process and is divided into three essential stores: sensory store (reception of environmental information for a few seconds), short-term memory (temporarily store that analyses, deconstructs and synthesizes information), and long-term memory (responsible for creating and saving mental constructs or schemas) [17]. Cognitive load theory provides guidelines on how to foster information retrieval and learning by enhancing either germane cognitive load by using standardized or well-known formats or by enhancing extraneous cognitive load by using the right visualization formats and designs [18].

For enhancing the extraneous cognitive load cognitive fit theory is important. It states that visualization always needs to fit the task at hand because the problem solver does not need to exert additional cognitive effort to either transform the problem representation to better match the task or to transform their decision processes to better match problem representation [30]. However, as mentioned in the introduction this theory produced contradicting results, indicating that other influences might be relevant as moderator and mediator variables.

Based on an extensive literature review (initial search: 1.952 articles; after abstract sorting based on the content: 237 articles) four essential dimensions of information visualization could be identified:

- Visual Complexity: Visual complexity is the degree of difficulty to transform an image into a consistent verbal description of the designer and the recipient of the visualization [32]. Ziemkiewicz and Kosara [33] claim that telling "meaningful stories is the goal of visualization…but stories can rise up purely from differences in shape and arrangement." Two components determine visual complexity, namely the used visualization type and the design or structure of a given visualization [19] [14] [15].
- Task Complexity: Task complexity has three variables: task type, task difficulty (or complexity) and task environment [27]. For task type two essential classifications could be found: first spatial (determining relationships and making comparisons) and symbolic tasks (use of discrete data values) by Vessey [30], and second accumulation (acquiring and recalling a single information cue), recognition (recognizing patterns and relationships between 2-3 information cues), estimation (identifying trends between numerous information cues), and projection (making projections to future values) by Hard and Vanecek [12]. Task difficulty can either be calculated objectively or tested subjectively by asking participants, and task environment represents time constraints, task interruptions etc. [28].
- Data Complexity: Data complexity combines data type and data density. For data type the accompanied dimensions need to be considered [16] and for data density the amount of data compressed in a visualization [11].
- Individual Complexity: Individual complexity can be clustered into three fundamental dimensions: Cognitive traits representing a persons' working memory ability, cognitive states representing situational and emotional influences and experience and biases [21]. Situational and emotional states revere to a persons' origin and current state. Of particular interest in conjunction with visual representations are knowledge and expertise, experience, decision making style, gender, and culture for the origin part as well as motivation, concentration and emotional issues that might focus attention [1] [13] [17] [18]. From the perspective of this research driven by cognitive load theory, prior knowledge, experience and expertise do have a high influence on information processing in terms of efficiency and effectivity [17].

An overview of the model as well as the possible dependencies, interdependencies and independencies can be seen in figure 1 (red framed are the dimensions are varied in this study to evaluate relationships, while the grey shaded ones were held constant. Later tests will vary).

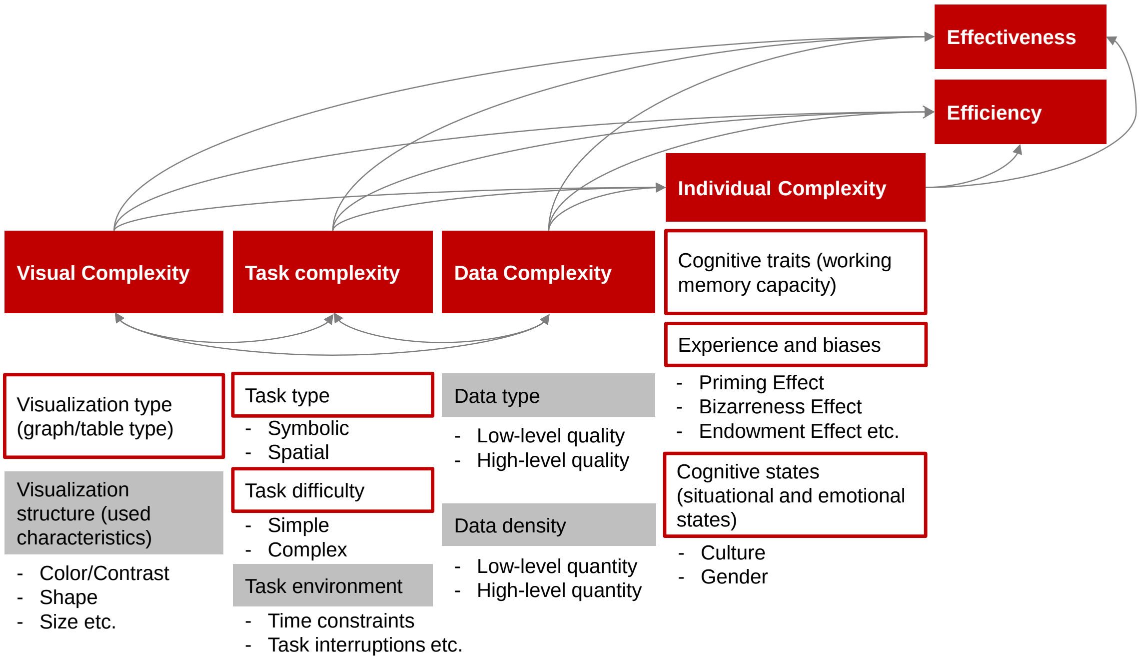

Figure 1. Data Visualization Score (DVS) Model

3 RESEARCH QUESTIONS

Despite early individual studies on correlations, there is a lack of research on a more holistic level, bringing these together into a unified model to predict the quality of visualization for managerial decision making. Therefore the following research question arises: How can a model be created to predict the goodness of a visualization for a certain task at hand considering a specific target audience? The variable TC was objectively measured by looking at behavioral acts and information cues caused by the used information visualization and its data density. The authors then came up with a preliminary model based on the main dimensions as laid out before. In a first step this model is now transformed into structural equations to cater for the interdependencies; and an early validation test is reported in this paper. The hypotheses we are looking out for in particular are:

H1: Does a lower task complexity (TC) improve Efficiency in cognition?

H2: Does a lower task complexity (TC) influence the correct interpretation of the data (Effectivity)?

4 METHODOLOGY

To operationalize the underlying variables, a mix of quantitative data collection methods is used, amongst: eye tracking experiments to gain insights into the scanning strategies and to analyze areas of interest in more detail, a validated, literature based questionnaire on cultural dimensions, an automated working memory span task to gain insight into the impact of working memory capacity on the ability to interpret visual information and several scores generated to assure all identified possible influences are either recorded or controlled by the experimental design. The experiment is designed as a double-random, between-group study and participants are assigned randomly to one of the groups. Participants engage in the experiment by performing four tasks of various complexities upon four varying visualization formats. Aim-oriented tasks are used because, based on observations, most people view graphs in order to obtain specific information [15]. Task types are used according to the classifications based on the cognitive fit theory [30] and in congruence with Hard and Vanecek [12], as those are the most cited classifications. Task complexity is calculated according to Wood [31] and is dependent on the visualization type used because it is predetermined by behavioral acts and information cues accompanied. Visualizations are designed and measured in accordance with literature and based on results of prior studies.

As mentioned eye tracking methodology is applied in this experiment. The experiment is conducted in a laboratory setting with consistent (artificial) light. Additional noise and other possible distractions are minimized. A remote eye tracker (SMI RED500 with a sample rate of 120 Hz), a nine point calibration, and, to ensure a minimum head movement during the recording, a headrest are used. After a maximum of ten minutes a recalibration is conducted to ensure a high quality data collection.

Participants of this initial validation trial are 23 students from the University of Applied Sciences Upper Austria. Task performance comparison between different participants and different experimental conditions is used to determine the effect of the independent on the dependent variables, creating 100 individual impressions for the calculation of the path weights.

The following table summarizes the variables from the framework of information visualization and their operationalization. Due to space limitations in this paper, the full variable specifications can be downloaded from the authors upon request.

Table 1. Overview on the variables, their operationalization and statistics

Variable	Measure	Operationalization and Statistics
Subjective visual complexity	Subjective visual complexity score	Assessment of the different visualization types by the participants (Likert Scale)
Task complexity	Task complexity score	Measure for objective task complexity by Wood [31] and different task types by Hard and Vanecek [12]
Data Complexity	Data complexity score	Quantity of information
Individual Complexity		
Cognitive traits	Working memory capacity	Automated working memory span tasks (based on E-Prime) by Redick et al. [23]
Cognitive states	Assessment based on questions	Questionnaire on cultural dimensions based on Hofstede (http://www.geerthofstede.nl/vsm-08)
Experience and biases	Assessment based on questions	Questions
Effectivity	Task accuracy	Answers to tasks
Efficiency	Task completion time; Scanning Strategies (sequential order strings)	Eye Tracking Measures; Time

In order to estimate the parameters of the underlying conceptual model, the authors turn to structural equation modeling (SEM) - in particular to variance-based SEM because of its mild distributional assumptions [24] and fewer convergence problems. The authors apply a version of partial least square path modelling because of its widely demonstrated capabilities of approximating latent variables using linear composites of observed variables.

As this model is at a relatively early stage, the authors want to be able not only to compute the path-coefficients, construct correlations and indicator weights, but also to derive an overall goodness-of-fit to compare with later, refined models. In order to do so, the authors thus transcend the classic PLS computation by using consistent and asymptotically normal PLS estimators for their equations [8]. This computation was supported by the software Smart PLS [25].

5 PRELIMINARY RESULTS AND DISCUSSION

The results of the study are a first step in order to validate the developed model. The endogenous variables and the weights of the formative indicators are displayed underneath.

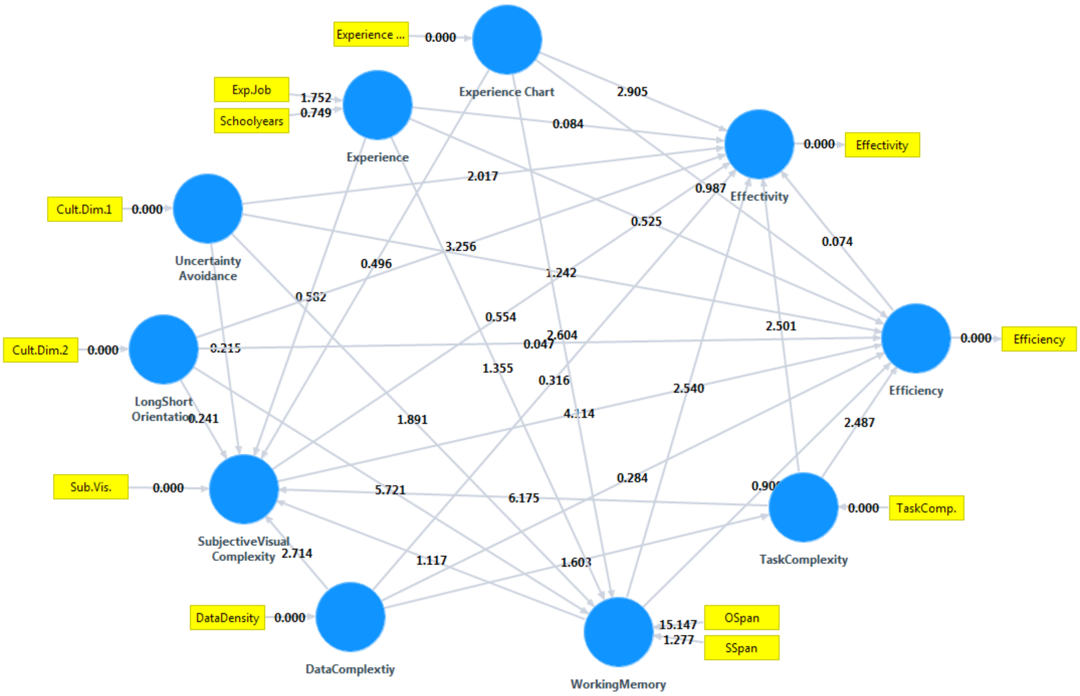

Figure 2. Model for DVS (Consistent PLS after Bootstrapping 1000 subsamples), significant if T score>1,96

Path Coefficients	OS	T Statistics	p-values
TaskComplexity => SubjectiveVisualComplexity	0,55	6,18**	0,00
LongShortOrientation => WorkingMemory	-0,47	5,72**	0,00
SubjectiveVisualComplexity => Efficiency	0,42	4,11**	0,00
LongShortOrientation => Effectivity	0,35	3,26**	0,00
ExperienceChart => Effectivity	-0,29	2,91**	0,00
DataComplexity => SubjectiveVisualComplexity	0,20	2,71**	0,01
ExperienceChart => WorkingMemory	0,21	2,60**	0,01
WorkingMemory => Effectivity	0,33	2,54**	0,01
TaskComplexity => Effectivity	-0,38	2,50**	0,01
TaskComplexity => Efficiency	0,26	2,49**	0,01
UncertaintyAvoidance => Effectivity	-0,17	2,02*	0,04
UncertaintyAvoidance => WorkingMemory	0,16	1,89°	0,06
DataComplexity => TaskComplexity	0,11	1,60	0,11
Experience => WorkingMemory	0,25	1,36	0,18
UncertaintyAvoidance => Efficiency	0,11	1,24	0,21
WorkingMemory => SubjectiveVisualComplexity	0,12	1,12	0,26
ExperienceChart => Efficiency	-0,08	0,99	0,32
WorkingMemroy => Efficiency	0,10	0,90	0,37

Table 2. path coefficients statistics sorted by p, excerpt through p<0,4.** significant p<0,01, * p<0,05, °p<0,1

6 INTERPRETATION AND CONCLUSION

H1 asked whether a low task complexity (TC) improves efficiency in cognition and H2 for its influence on the effectivity (correct interpretation). As we can see in table 2, TC predicts Efficiency and Effectivity at a significance level of p=0.01. Therefore we can confirm both hypotheses. Overall adjusted model R^2 for Efficiency is 34.3% and for Effectivity it is 33.3%.

Besides the confirmed factor TC, the main predictors for effectivity are cultural dimensions (long vs short-term orientation index and uncertainty avoidance index), experience with the presented chart-type and working memory capacity; while the main predictors for efficiency are the subjective visualization score and as confirmed, TC. To further validate the model, discriminant analyses, VIF tests for collinearity and F-tests were conducted and can be downloaded upon request. Overall, this preliminary study looks promising as it can be seen as an early endorsement of the created model and the DVS. The highly interlinked model allows the authors to create an optimal data visualization given a specific task for a specific audience, concerning experience and various cultural backgrounds. This may have great implications for example for improving standard reports to find an optimal efficient, audience and task adjusted, reporting design.

REFERENCES

[1] Al-Kassab, J., Ouertani, Z. M., Schiuma, G., Neely, A. (2014): Information visualization to support management decisions, In: International Journal of Information Technology & Decision Making, Vol. 13, No. 2, pp. 407-428

[2] Barat, A. H. (2007): Human perception and knowledge organization: visual imagery, In: Library Hi Tech, Vol. 25, No. 3, pp. 338-351

[3] Beattie, V., Jones, M. J. (2002): Measurement distortion of graphs in corporate reports: an experimental study, In: Accounting, Auditing & Accountability Journal, Vol. 15, No. 4, pp. 546-564

[4] Cho, C. H., Michelon, G., Patten, D. M. (2012): Impression management in sustainability reports: an empiri-cal investitgation of the use of graphs. In: Accounting and the Public Interest, Vol. 12, pp. 16-37

[5] Conati, C. and Maclaren, H. (2008): Exploring the role of individual differences in information visualization, In: Advanced visual Interfaces (Proceedings of AVI'08), ACM, pp. 199-206

[6] Dilla, W. N., Janvrin, D. J. (2010): Voluntary disclosure in annual reports: the association between magnitude and direction of change in corporate financial performance and graph use. In: Accounting Horizons, Vol. 24, No. 2, pp. 257-278

[7] Dilla, W., Janvrin, D. J., and Raschke, R. (2010): Interactive data visualization: new directions for accounting information systems research, In: Journal of Information Systems, Vol. 24, No. 2, pp. 1-37

[8] Dijkstra, T.K., Henseler, J., 2015. Consistent and asymptotically normal PLS estimators for linear structural equations. Computational Statistics and Data Analysis. 81 (2015) 10-23

[9] Falschlunger, L., Eisl, C., Losbichler, H., Greil, A. (2014): Improving information perception of graphical displays – an experimental study on the display of column graphs, In: Full Paper Proceedings of the WSCG 2014 (Conference on Computer Graphics, Visualization and Computer Vision), 2nd -5th June 2014, pp. 19-26

[10] Galletta, D., and Vessey, I. (1991): Cognitive fit: an empirical study of information acquisition, In: Information Systems Research, Vol. 2, No. 1, pp. 63-84

[11] Gelman, A., Unwin, A. (2013): InfoVis and statistical graphics: different goals, different looks, In: Journal of Computational & Graphical Statistics, Vol. 22, No. 1, pp. 2-28

[12] Hard, N. J., and Vanecek, M. T. (1991): The implications of task and format on the use of financial information, In: Journal of Information Systems, Vol. 5, No. 2, pp. 35-49

[13] Hahn, U. (2014): Experiential limitation in judgment and decision, In: Topics in Cognitive Science, Vol. 6, No. 2, pp. 229-244

[14] Hill, W. Y., and Milner, M. M. (2003): Guidelines for graphical displays in financial reporting, In: Accounting Education, Vol. 12, No. 2, pp. 135-157

[15] Huang, W., and Eades, P. (2005): How people read graphs, In: Proceedings of the APVIS 2005 (Asia Pacific Symposium on Information Visualization – Conference in Research and Practice in Information Technology), Sydney, Vol. 45, pp. 51-58
[16] Kuang, X., Zhang, H. Zhao, S., McGuffin, M.J. (2012) Tracing tuples across dimensions: a comparison of scatterplots and paralell coordinate plots, In: Computer Graphics Forum, Vol. 31, No. 3, pp. 1365-1374
[17] Lord, R. G., and Maher, K. J. (1990): Alternative information-processing models and their implications for theory, research, and practice, In: Academy of Management Review, Vol. 15, No. 1, pp. 9-28
[18] Mostyn, G. R. (2012): Cognitive Load Theory: What it is, why it's important for accounting instruction and research; In: Issues in Accounting Education, Vol. 27, No. 1, pp. 227-245
[19] Parsons, P., and Sedig, K. (2014): Adjustable properties of visual representations: improving the quality of human-information interaction, In: Journal of the Association for Information Science & Technology, Vol. 65, No. 3, pp. 455-482
[20] Parush, A., Hod, A., Shtub, A. (2007): Impact of visualization type and contextual factors on performance with enterprise resource planning systems, In: Computers & Industrial Engineering, Vol. 52, No. 1, pp. 133-142
[21] Peck, E. M., Yuksel, B. R., Harrison, L., Ottley, A., and Remco, C. (2012): Position paper: towards a 3-dimensional model of individual cognitive differences, In: Beyond Time & Errors – Novel Evaluation Methods for Visualization (Proceedings of BELIV 2012), ACM, pp. 1-6
[22] Porat, T., Oron-Gilad, T., and Meyer, J. (2009): Task-dependent processing of tables and graphs, In: Behaviour & Information Technology, Vol. 28, No. 3, pp. 293-307
[23] Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, N. U., Kane, M. J., and Engle, R. W. (2012): Measuring working memory capacity with automated complex span tasks, In: European Journal of Psychological Assessment, Vol. 28, No. 3, pp. 164-171
[24] Reinartz,W., Haenlein, M., Henseler, J.,2009. An empirical comparison of efficacy of covariance-based and variance-based SEM. Int. J. of Research in Marketing. 26 (4), 332-344
[25] Ringle, C. M., Wende, S., and Becker, J.-M. 2015. "SmartPLS 3." Boenningstedt: SmartPLS GmbH, http://www.smartpls.com.
[26] So, S., and Smith, M. (2004): Multivariate decision accuracy and the presentation of accounting information, In: Accounting Forum, Vol. 28, No. 3, pp. 283-305
[27] Speier, C., Vessey, I., and Valacich, J. S. (2003): The effects of interruptions, task complexity, and computer-supported decision-making performance, In: Decision Sciences, Vol. 34, No. 4, pp. 771-797
[28] Speier, C. (2006): The influence of information presentation formats on complex task decision-making performance, In: International Journal of Human-Computer Studies, Vol. 64, No. 11, pp. 1115-1131
[29] Toker, D., Conati, C., Steichen, B., and Carenini, G. (2013): Individual user characteristics and information visualization: connecting the dots through eye tracking, In: Proceedings of the CHI 2013, April 27th to May 2nd, Paris, pp. 295-304
[30] Vessey, Iris (1991): Cognitive fit: a theory-based analysis of the graphs versus tables literature, In: Decision Science, Vol. 22, No. 2, pp. 219-240
[31] Wood, R. E. (1986): Task complexity: definition of the construct, In: Organizational Behavior and Human Decision Processes, Vol. 37, pp. 60-82
[32] Yigitbasioglu, O. M., and Valcu, O. (2012): A review of dashboards in performance management: implications for design and research, In: International Journal of Accounting Information Systems, Vol. 13, No. 1, pp. 41-59
[33] Ziemkiewicz, C, and Kosara, R. (2010): Beyond bertin: seeing the forest despite the trees, In: IEEE Computer Graphics & Application, Vol. 30, No. 5, pp. 7-11.