Output Based upon Input, the Result from Data Analytics and Visualization of TNI Registration System's Data

Prajak Chertchom, Ph.D.

Faculty of Information Technology, Thai-Nichi Institute of Technology
1771/1 Pattanakarn Rd., Suanluang, Bangkok 10250, Thailand

prajak@tni.ac.th

Abstract - **The large amount of computerized data in business has been steadily increasing over the years. Data extraction using data analytics and visualization have been widely used to discover information knowledge from various areas. This paper analysed data with statistic tools and visual analysis to extract and gather information from registration system in order to;**

1. Understand students and predict an academic performance condition.

2. Compare a knowledge's result from statistic tools with graphical output from visualization software.

Registration data was obtained from department of academic affair of Thai-Nichi Institute of Technology from year 2007-2015. In this study, we only focused on IT student's data that contains 1,822 rows. Based on the applied techniques that we have developed, we found that academic performance (GPA) of TNI's IT students have lower average GPAs and not different result from their previous level (High school). In addition, data analytic and visual analysis can help a university to evaluate their teaching performance, gain insight into their students, discover knowledge and make evidence-based for future educational planning decisions. Moreover, Data visualization allows an organization to make better business decisions from allows insights into a big data.

Keywords - **Data analytic, Visual Analysis, regression analysis, academic performance, institutional strategic plan**

I. INTRODUCTION

In the business world today, the enterprises' data is recorded in a digital format. This data is increasing at an incredible rate. Data size also becomes too big and complex to analyse. Several organizations have a very big data in their storage; however, they do not use this data for understanding their customer, employees, operations as well as stakeholders. It is not only enterprises who are now facing this problem, universities also need to learn what they are operating within their business, analyse their educational process, react to students' academic performance, and make decision that will help manage to produce a good output to an industry.

Every semester, universities collect massive amounts of data internally and externally as they interact with students.

Thai-Nichi Institute of Technology (TNI) is one of private university in Thailand that is using the student registration system from starting operation for course registration and a central part of the educational administration system for undergraduate and graduate students. Since 2007, TNI's registration system has stored student's profiles both demographic characteristics and grade point average value for a course, created many transaction files and accumulated data for almost 10,000 rows including student's profile and grade. This system is just used for enabling student to register online and supporting all stakeholders for normal administrative function. Despite the fact that there are a lot of questions within university's operations, for example, do we deliver a proper student to industry, how do we recruit more students next semester, our curriculums are suit with today business.

For a very high competitive in educational business, if TNI stop doing self-analysis for business planning and developing its educational services, TNI may face a problem in the near future. Thus, data analytics and virtualization can transform raw fact and numeric into strategic insights that deliver the understanding of TNI's business and use it to solve specific and complex educational problems on reliable information, not gut-instinct. Moreover, it can equip TNI to make fact-based decisions across value chain to increase overall performance, differentiation and profitable growth.

Thus, this study aims to find out new opportunities and knowledge from data analytic and visualization for developing TNI's curriculum, understanding students and findings ideas for institutional strategic planner to support their forecasting and resource planning from data analytics and visualization.

In addition, this study contributes to the area of a research study in the field of academic analytics and related to the field of educational data mining and learning analytics. [1]

II. THEORETICAL BACKGROUND AND FRAMEWORK

A. Approaches to institutional strategic plan

Data analysis is a tool for helping organizations to understand their customers, business activity in order to use for marketing planning, self-analysing and forecasting what might happen in the future. The use of data analysis is also used to categorize profile and describe the demographic of customers for understanding what happened to current business situation. [2]

According to steadily increasing of the large amount of computerized data in business as well as in educational sector over the years, data extraction using data analytics

and visualization have been widely used to discover information knowledge from various areas. [3] This paper used data analytics with statistical software and visualization techniques to extract and gather information from registration system in order to understand students, TNI's business and predict academic performance condition information.

In addition, the increasing of university in Thailand has reached 155 institutions (Government = 83, private = 72) [4], Thai population birth rate has been decreasing in opposite. From the report of National Statistical Office Thailand (2015) [5], young population is largely decline from 27.9 percent in1995 to 23.0 percent in 2005. As the result, this changing of the age structure would affect to educational business in term of finding new students, strengthen teaching system and curriculum preparation to attract parents. As summary, TNI needs to do data analytic to cope this issue.

B. Data Analytics

According to Martin (as cited in Grant et al., 2010) [6], data analysis is a process to collect, organize, validate and classify of the entities, attributes ad relationship to understand an information that exist. Sherman [7] also noted that data analytic is a data designing techniques that help businesses reporting, analytics, predictive modelling, and decision-making. In addition, from the academic analytics [8], they propose the data analytic engine that consists of five steps: capture, report, predict, act, and refine. To turn data into "actionable" information, a guideline from previous studies have been reviewed, and applied those techniques to be a new analytic process flow as shown in figure 1.

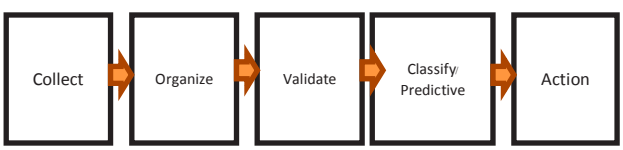

Fig. 1 Analytic process flow (by author)

The first is an activity to collect and/or retrieve data, results, context and other transactions. It is important to understand the structure or logical concept of data, and the type of information that gather from application/system. The second task is to organize data in ways that make them easier to work with for example in this study we prepare them in Microsoft Excel and use Microsoft Access to create linkage with each data file.

The second step is an activity to transform data that was collected in forms other than numbers, and turned into quantitative data for analysis. For instance, qualitative data such as family status, birthplace will be changed into numbers, by assigning number to quantitative value.

The third step is the process of data validation to ensuring that a study's data are clean, correct and computable. Although the data of this research came from TNI's registration system, some data input by students were unable to use for statistical computation and files were separated and not in one piece. This study used "validation rules" to check for correctness, meaningfulness and used Microsoft Access to join file and do a data type checks. In addition, we used cross-system consistency checks to compares data from different files to ensure it is consistent. Moreover, we transformed data to a common format to be compared.

Then, we analyse these data by classification and predictive techniques using data visualization and statistical computation with regression analysis.

C. Relationship modelling with Regression analysis

Many scholars apply multiple regression models for the final stage of data analysis. One of benefits of regression analysis is, for instance, it can predict the expected GPAs based on combinations of variables as they may be configured in the backgrounds of individuals (e.g., parents, habits, and birthplace). In this study of GPAs, a regression model can tell the strength of the factors in predicting academic success, predict factors that affecting the differences among students. In addition, one of the advantages of multiple regression models is that they allow researcher to document the influence of independent variables on GPAs and consider the relationship between them to suggest about theories that posit innate differences in abilities to succeed in university. [9], [10], [11] In the case of this study, after we clean and validate the data, we can use multiple factors to be a regression model to identify the factors that are most responsible for academic success as shown in figure 2.

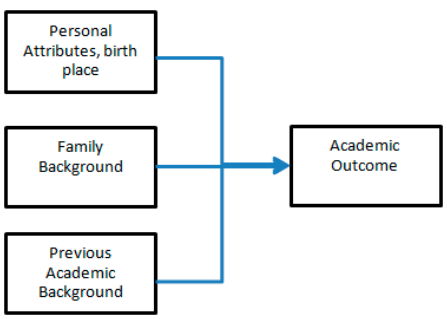

Fig. 2 Regression analysis model (by author)

D. Data Visualization Analysis

Data analytics with visual analysis have been widely used to discover information knowledge from various areas. Visualization techniques is a key tool for today business that help management in clearly understanding the complicated data and support decision making. [12], [13], [14] Visual analysis also provides a new way to explore large data sets to "detect the unexpected" [15]

Among several analytics tools, data visualization helps different users to explore and make sense of their data. Adding visual analytics to decision-making process is also helping uncover insights buried in data. In addition, comparing to others tools, analytics visualization helps organization quickly recognize outliers that may affect service quality or customer churn. Some of statistical analysis such as descriptive and correlation will not be able to give obvious information, but visual analytics support a capability that can explain the relationships between variables with graphical technologies provides an understanding of geospatial data. [15], [16]

Thus, from the theory background and problems statements, this study applied the two major of analytics deployment which are regression analysis and data visual analysis for discovering a knowledge that affects to student performance and monitoring a related background that engages to student learning.

III. RESEARCH METHODOLOGY

A. Data collection

The registration data was obtained from department of academic affair of Thai-Nichi Institute of Technology from year 2007-2015. In this study, we only focused on IT student's data and the total number of students in the study was 1,822 rows. The data types that were selected are students' demographic, academic ability (GPA), historic students and family background. We use Microsoft Excel and Microsoft Access to organize the data for analytic and visualization process.

B. Data Structure

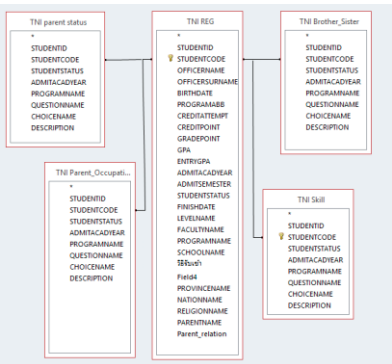

Fig. 3 Data Structure of TNI's registration data

The data that obtained from TNI's registration are separately in different file structure. Thus we organized data using MS Access to suit this study's specific purposes as shown in figure 3.

C. Measurement

The analysis of data in this study was applied the five steps of academic analytics [8] which are capture, report, predict, act, and refine and designed new analytic process flow to provide useful insights into the learning process of students' data. (As shown in Figure 1) In addition, the extraction of data also used both descriptive and inferential statistics (using SPSS) that is a regression analysis to analyse and predict the large dataset and compare a result with graphic data from visualization software. (Rapid Miner Studio Basic and Micro Soft Excel)

IV. EMPIRICAL STUDY

A. Descriptive Statistics

TABLE I:
Result from Univariate Statistics

N = 1,774	GPA	Entry GPA
Mean	2.29	2.83
Standard Deviation	.969	.498
Other Factors		
Average number of Brothers and Sisters	2.08	
Average Family Income	415,925.61	

*N= 1,774 (after cleansing data)

TABLE II:
Crosstab of no of student and GPA by year

Year	<2.00		2.00-2.50		2.60-3.00		3.00 and above	
2007	4	12.10%	10	30.30%	10	30.30%	9	27.30%
2008	22	18.00%	30	24.60%	33	27.00%	37	30.30%
2009	27	15.80%	33	19.30%	62	36.30%	49	28.70%
2010	54	24.20%	61	27.40%	63	28.30%	45	20.20%
2011	65	25.50%	60	23.50%	76	29.80%	54	21.20%
2012	85	39.40%	50	23.10%	37	17.10%	44	20.40%
2013	59	27.20%	62	28.60%	46	21.20%	50	23.00%
2014	102	35.70%	62	21.70%	58	20.30%	64	22.40%
2015	98	39.00%	57	22.70%	53	21.10%	43	17.10%
Total	516	29.10%	425	24.00%	438	24.70%	395	22.30%

TABLE III:
Frequency Table –Student by Region

Year	BKK	Central (exclude BKK)	N	S	NE	E	W
2007	15	5	1	1	8	1	1
2008	83	11	3	6	9	4	5
2009	127	22	2	8	6	4	1
2010	159	28	3	13	10	7	2
2011	186	25	7	8	15	6	8
2012	145	37	2	4	18	7	2
2013	151	40	2	11	4	4	5
2014	183	46	3	14	12	20	7
2015	152	52	9	8	3	10	9
Total	1,201	266	32	73	85	63	40

BKK: Bangkok N: North, S: South, NE: North East, E: East, W: West

TABLE IV:
Frequency Table –Student by Region *(%)*

Year	BKK	Central (exclude BKK)	N	S	NE	E	W
2007	45.5 %	15.2%	3.0%	3.0%	24.2 %	3.0%	3%
2008	68.0 %	9.0%	2.5%	4.9%	7.4%	3.3%	4%
2009	74.3 %	12.9%	1.2%	4.7%	3.5%	2.3%	1%
2010	71.3 %	12.6%	1.3%	5.8%	4.5%	3.1%	1%
2011	72.9 %	9.8%	2.7%	3.1%	5.9%	2.4%	3%
2012	67.1 %	17.1%	0.9%	1.9%	8.3%	3.2%	1%
2013	69.6 %	18.4%	0.9%	5.1%	1.8%	1.8%	2%
2014	64.0 %	16.1%	1.0%	4.9%	4.2%	7.0%	2%
2015	63.3 %	20.7%	3.6%	3.2%	1.2%	4.0%	4%
Total	68.1 %	15.0%	1.8%	4.1%	4.8%	3.6%	2%

BKK: Bangkok N: North, S: South, NE: North East, E: East, W: West

A crosstab table 2 and 3 simply present the absolute frequency broken down by categories of number of students with GPA and categories of number of students with regions that they are from respectively. It is also possible to find percentages in these types of tables as shown in table 3.

These tables let us know how many times a particular category occurs. Table 1 and Table 2 show that most of IT students fall into the average GPA with lower than 2.00 and between 2.00 and 2.50. In addition, table 3 and 4 show that most of IT students are from Bangkok area with 68.1% from all. Moreover, table 1 presents average number of brothers and sisters equals 2.08 and average family income equals 415,925.61

B. Use Regression Analysis to Determine Validity of Relationships

In this study, we also used regression analysis to determine of the relationship between variables to shows what the value of Y (GPA) would be if X (independent variables) were changed. In this case, we want to identify the factors that are most responsible for academic performance (GPA).

TABLE V:
Summary output by regression analysis

Regression Statistics	
Multiple R	.403
R Square	.163
Adjust R Square	.157
Standard Error	.873
Observations	1,034

TABLE VI:
Output by regression analysis

Predictor	Coef	StDev	T	Sig.
constant	96.800	26.770	3.616	.000
Entry GPA	.795	.049	16.105	.000
Academic Year	-.038	.010	-3.672	.000
Enter Channel	.005	.015	.297	.766
Region	.004	.015	.263	.792
High School/Vocational	.184	.031	6.011	.000
Family status	-.037	.055	-.017	.493
Number of Brothers/Sisters	.044	.026	.040	.097

From table 5 and 6, the regression equation comprising the explanatory variables is

GPA = 96.8 + .795 Entry GPA + -.038 Academic Year + .005 Enter Channel + .004 Region + .184 High School/Vocational + -.037 High School/Vocational + .044 Number of Brothers/Sisters

This analysis reveals that GPA of TNI IT Students depends on previous academic performance. The t-ratio for Entry GPA is highly significant. The conclusion to be drawn is that there is evidence of a relationship between GPA and previous academic result, the others factors such as home and community environment (region where students are from) and quality of family have generally lower effected to students' academic performance at TNI.

C. Data Visualization Analysis

This study also used visualization analysis to present the data with graphical format. At the conclusion, we want to conduct an experiment the two way of analysis how statistic or graphical tools give a decision quality.

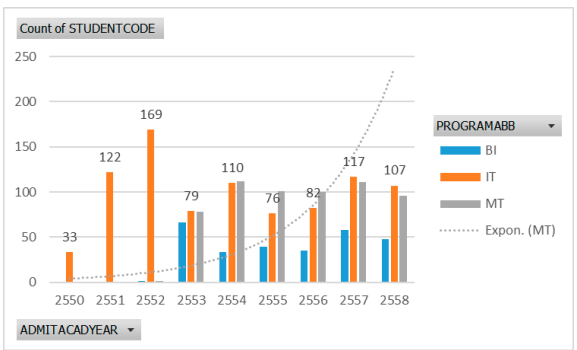

Fig.4 Number of Students by Information Technology Faculty's Program

The bar chart shown in Figure 4 shows number of students by information technology faculty's Program (IT). It reveals that the biggest proportion of this faculty's program is IT with total 895.

The trend line also implies that total number of IT students for all majors at TNI has been slightly increasing recently.

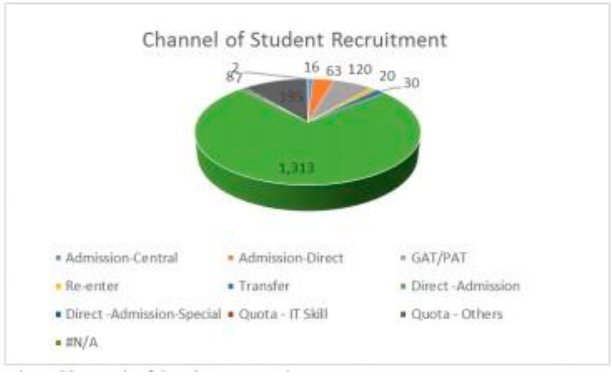

Fig.5 Channel of Student Recruitment

Figure 5 shows that most of students entered to TNI's IT faculty by the channel of direct admission with 1,313 students and by quota with 195 students respectively.

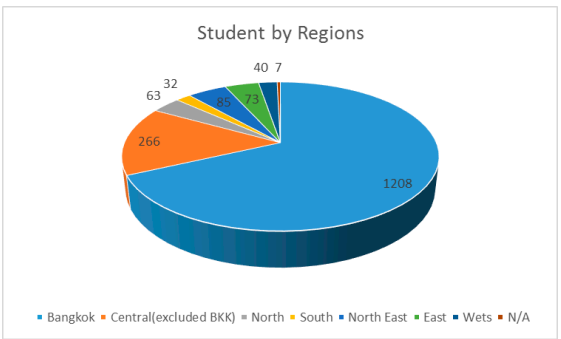

Fig.6 Student by Regions

Figure 6 shows that most of students entered to TNI's IT faculty are from Bangkok area and city not far from Bangkok with number 1,208 and 266 respectively.

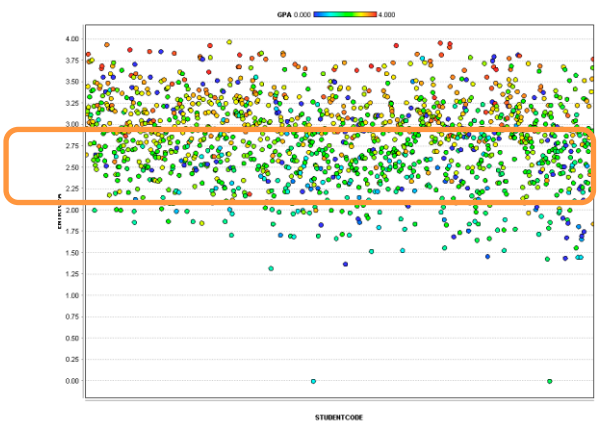

Fig.7 Scatter of Entry GPA

The scatter plot in figure 7 shows entry GPA of each student from 2008 – 2015. The green density of plot describes that most of GPA are between 2.00 and 2.75 and this picture reveals that TNI has recruited a not high academic performance student from high school.

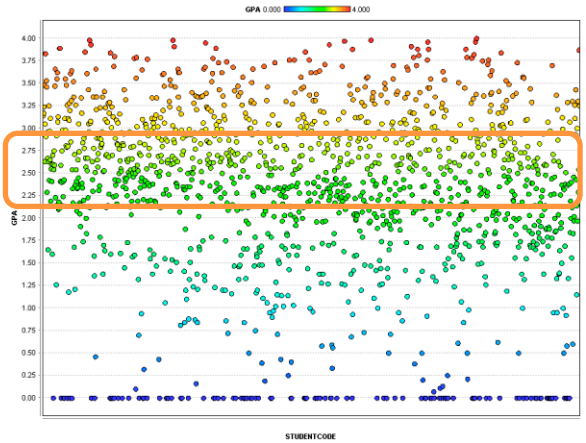

Fig.8 Scatter of Graduated GPA

The scatter plot in figure 8 shows GPA of graduated students from 2008 – 2015. The green density of plot describes that most of GPA are between 2.00 and 2.50 and this picture reveals that students graduated with a not high academic performance and quite close to their previous result from high school.

V. DISCUSSION AND FUTURE WORK

Registration systems has accumulated amounts of data every semester. Without a good data analytic and visualization analysis, this data should be worthless. There are many tools for being applied to help make sense of this data. In this study, we proposed two tools: statistical analysis and visualize analytics to help us thrive in strategic teaching plan.

From output in previous section, we can summarize that:

1. The outputs from two tools gave the same result in different format and perspective. However, each tools provide the result with different information.

2. Descriptive analytics is used to summarize collection of information that tell us what happened from past to present. It provides simple summaries that is still difficult to apply it for decision-making.

3. Predictive analytics is the next step up in this study with regression analysis. This technique is learning techniques to study recent and historical data, that allowing us for predicting about the future. However, the weak point is that the prediction is based on significant that still not precisely understanding and predictive analytics are probabilistic in nature; it can just only tell us what might occur in the near future.

According to Hirsch et al., with implications of the cognitive fit model, from those computation techniques, the one who uses that information needs to translate a numbering format to a form for more precisely understanding. For easy to solve the problems, we should visualize data with graphs as depicted in figure 1 to 5. [15]

In addition, the most distinctions between statistics analytics and visualize analytic, visualized graph can deliver knowledge and communicate to receiver easier than number system.

Finally, from author's opinion from this experiential research, if we are plan about the future scenarios with

various intersections of decision, we have to combine the two tools to explore the new future scenario visual and to explore data to **"**detect the unexpected**"** as recommended in table 7**:**

TABLE VII**:**
Why to choose analytic tool **(**By Author**)**

	Statistical technique		
	Descriptive	**Inferential**	**Visual Analytic**
Ease of Use	●	◑	◑
Ease of communication	◑		●
Visual Analytic			●
Tell What Happen			●
Prediction		●	◑
Business User focus	◑	◑	●

REFERENCES

[1] I. Ognjanovic, D. Gasevic, and S. Dawson, "Using institutional data to predict student course selections in higher education," *Internet and Higher Education*, vol. 29, pp. 49–62, 2016.

[2] AnneMarie Scarisbrick-Hauser, "Data analysis and profiling," *Direct Marketing: An Intl Jnl*, vol. 1, no. 2, pp. 114–116, Jun. 2007.

[3] C. Vialardi *et al.*, "A Data Mining Approach to Guide Students Through the Enrollment Process Based on Academic Perfor mance," *User Modeling and User-Adapted Interaction*, vol. 21, no. 1–2, pp. 217–248, Apr. 2011.

[4] Office of Higher Education Commission (HEC). (2016, Mar 1). *University in Thailand* [Online] Available: http://www.mua.go .th/university.html

[5] National Statistical Office of Thailand. (2016, Mar 5). *Thailand Population* [Online] Available: www. nso.go.th

[6] G. Kevin, H. Ray, and E. David, *Strategic Information Systems Management*, 1st ed. Andover: Cengage Learning EMEA, 2009.

[7] R. Sherman, "Chapter 15 - Advanced Analytics," in *Business Intelligence Guidebook*, R. Sherman, Ed. Boston: Morgan Kaufmann, 2015, pp. 375 – 402.

[8] C. John P., D. Peter B., and O. Diana G., "Academic Analytics: A New Tool for a New Era," *EDUCAUSE Review*, vol. 42, no. 4, p. 40, 2007.

[9] A. O. Julian, Q. Sally, and K. Rachel, "Attachment style, social skills, and Facebook use amongst adults," *Computers in Human Behavior*, vol. 29, no. 3, pp. 1142–1149, May 2013.

[10] M. Michikyan, K. Subrahmanyam, and J. Dennis, "Facebook use and academic performance among college students: A mixed-methods study with a multi-ethnic sample," *Computers in Human Behavior*, vol. 45, pp. 265–272, Apr. 2015.

[11] S. Ainin, M. M. Naqshbandi, S. Moghavvemi, and N. I. Jaafar, " Facebook usage, socialization and academic performance," *Computers & Education*, vol. 83, pp. 64 – 73, 2015.

[12] K. Platts and K. H. Tan, " Strategy visualisation: knowing, understanding, and formulating," *Management Decision*, vol. 42, no. 5, pp. 667–676, 2004.

[13] C.-H. Kao, J. Nakano, S.-H. Shieh, Y.-J. Tien, H.-M. Wu, C. Yang, and C. Chen, "Exploratory data analysis of interval-valued symbolic data with matrix visualization," *Computational Statistics & Data Analysis,* vol. 79, pp. 14 – 29, 2014.

[14] R. Ashman and A. Patterson, "Seeing the big picture in services marketing research: infographics, SEM and data visualisation," *Journal of Services Marketing*, vol. 29, no. 6/7, pp. 613–621, 2015.

[15] V. L. Lemieux, B. Gormly, and L. Rowledge, "Meeting Big Data challenges with visual analytics: The role of records management," *Records Management Journal*, vol. 24, no. 2, pp. 122–141, 2014.

[16] A. A. Ganah, "The Use of Computer Visualisation Communica ting Constructability Information in UK," *Journal of Engineering, Design and Technology*, vol. 1, no. 2, pp. 151–167, 2003.

[17] B. Hirsch, A. Seubert, and M. Sohn, "Visualisation of data in management accounting reports: How supplementary graphs improve every-day management judgments," *Journal of Applied Accounting Research*, vol. 16, no. 2, pp. 221–239, 2015.