ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

A competing design for the income statement.

Name: Manon Commijs
Master: Accounting, Auditing & Control
Date: 30th of September 2017
Supervisor: Prof. dr. Maarten Pronk
Second reader: Drs. Rob van der Wal RA

Abstract

In this thesis I designed a competing design for the income statement in which I incorporate sparklines in the tabular format to show an overall trend of the financial data. With an experiment based on a between subjects design I answer the following research question: "Will the competing design for the income statement positively affect the decision making of users of the financial statements." The variables are accuracy and timing and task complexity is included as a mediating variable. The results do not support my expectation except for the accuracy with questions of low task complexity. Even after I eliminated four subjects the results did not alter. Therefore, I can't say that the competing design positively affects the decision making of users of the financial statement.

Contents

1 Introduction

The FASB noted in 2010, paragraph 43, the purpose of the financial statement presentation: "How an entity presents information in its financial statements is critical to effectively communicating that information to those outside the entity. Effective financial statement presentation provides disaggregated information organized in a manner that communicates clearly a cohesive financial picture of an entity." (IFRS Foundation, 2010). Financial information in the financial statement is primarily portrayed in a tabular format. However, it is also common financial reporting practice to use graphical formats to communicate corporate information (Hill & Milner, 2003). The graphical displays in financial reporting communicate the quantified corporate information. Thus, it is not included to portray the financial information, such as the income statement.

A graphical format and tabular format both offer different views. The graphical format provides a holistic view, it portrays the 'big picture' and emphasizes relationships in the data (Vessey, 1991). The tabular format, on the other hand, offers an analytical view (symbolic) and presents numerical values. Frownfelter-Lohrke (1998) and Lucas (1981) combined the graphical and tabular format (combination format) and found that there is some evidence that subjects receive the best information from the combination format in different types of tasks. Therefore, users of financial statements might benefit, make better decisions, from a graphical presentation in addition to the tabular statements.

In previous studies the competing design of the tabular format were standard graphical formats. So and Smith (2004) presented a tabular-graphical combination with a separate table and an additional bar graph. Desanctis and Jarvenpaa (1989) and Frownfelter-Lohrke (1998) also combined, but still showed separately, a tabular and graphical format. For the graphical format they used the horizontal bar charts. In a study by Parsons and Tinkelman (2013) another graphical format was incorporated in their combination format, known as sparklines. Sparklines is a graphical format designed by Edward Tufte which he introduced in 2006. He describes this graphical format as small, high-resolution graphics usually inserted in a full text of words, images and numbers (Tufte, 2006). Sparklines are useful for continuous data since they present a combined pattern, a general shape and plenty of detail. Therefore, for the competing design of the tabular format I incorporate the sparklines in the tabular format to show an overall trend of the financial data. This results in the following research question:

Research question: Will the competing design for the income statement positively affect the decision making of users of the financial statements.

Research design

To answer the research question I conduct an experiment based on a between-subjects design. For the experiment the subjects will adopt the role of investor for four companies. They take upon this role, because I am interested in seeing how well decision makers perform with different presentation formats.

The experiment consists of two treatments. In the first treatment (company A & B) the subjects receive the income statement of the (third) quarterly earnings announcement of the company in either tabular- or combination format. The accuracy and timing of the subjects is measured under high and low task complexity. In the second treatment (company C & D) a memorizing element is included and the accuracy of the subjects is measured. The subjects get 120 seconds to interpret the income statements of the (third) quarterly earnings announcement for both companies, one in the tabular format and the other in the combination format. The subjects, in total 40, of the experiment are university students in their masters. Specifically, most of the subjects are from the master Accounting, Auditing & Control.

Findings and Contribution

I expected that the accuracy of users would be better especially with high task complexity. I also expected that the decision quality of users would be better with the competing design instead of the tabular format. However, the results do not support my expectation except for questions of low task complexity. Even after I eliminated four subjects the results did not alter. Therefore, I can't say that the competing design positively affects the decision making of users of the financial statement.

Even though I did not find a statistical confirmation that the combination format positively affects the decision making of the users of the financial statement, I do make several contributions with this research. I found statistical evidence that there is no difference in the accuracy of the users of the financial statement between the two formats no matter the task complexity. Additionally I observed that the subjects are positive about sparklines. This is interesting for policy makers and companies since users of financial statements are positive about sparklines. It also doesn't have a negative impact on their decision making. Therefore, implementing the combination format won't result in less accurate decision making. However, further research regarding sparklines is recommended to determine the efficiency of the combination format and the implementation in the financial statement.

2 Literature review

In this chapter I discuss relevant literature for my thesis. My thesis relates to two streams of literature. First, it relates to literature on graphical reporting, this focuses more on the design of a graph or table. Second, it relates to management accounting literature on presentation format, graphical or tabular presentation. Both streams of literature take into account psychological literature about cognitive elements in presentation format and graphical reporting.

2.1 Graphical reporting

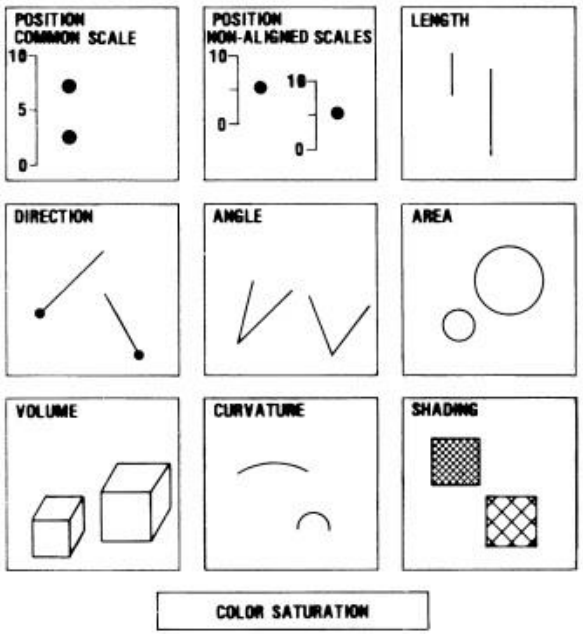
Graphical reporting is the practice of using graphical formats for communicating information. To be precise, it is common financial reporting practice to use graphical formats to communicate corporate information (Hill & Milner, 2003). The distinct feature of the graphical format is that they portray the 'big picture' and emphasize relationships in the data (Mackay & Villarreal, 1987; Vessey, 1991). Graphical formats are also eye catchers, they attract and direct the readers' attention. The design has a stronger impact compared to a tabular format (Davis, 1989). Essentially, graphical formats can improve the effectiveness and efficiency of communication and data analysis (Cleveland & McGill, 1984; Mackay & Villarreal, 1987; Hwang, 1995; Frownfelter-Lohrke, 1998). This is similar for the enhanced usefulness of information for business analysis and financial forecasting (Desanctis & Jarvenpaa, 1989).

Graphical formats can take many forms. The most common graphical format for presenting financial information in experimental settings is the bar chart (Davis, 1989; Amer, 1991; So & Smith, 2004; Cardinaels, 2008). Another popular graphical format for presenting financial information in experimental settings is the multivariate graphical display method, also known as the facial format, introduced by Chernoff (1973) (Stock & Watson, 1984; Mackay & Villarreal, 1987; Amer, 1991; So & Smith, 2004). In practice, the most commonly used standard graphs are the line graphs, bar charts and pie charts (Hill & Milner, 2003), however financial reporting might also include flow charts and mapping. Cleveland and McGill (1984) discuss a variety of common graphical formats: sample distribution function plot, bar charts, pie charts, divided bar charts, statistical maps with shading, curve-difference charts, cartesian graphs, triple scatterplots, volume charts and juxtaposed cartesian graphs.

Graphical communication and perception

Each graphical format presents data in a different way. To extract information from a graphical format a reader uses an elementary perceptual task. The term elementary perceptual task means that a reader needs to perform one or more mental visual tasks to extract values of variables from a graph (Cleveland & McGill, 1984). Cleveland and McGill (1984) describe 10 decoding tasks, see figure 1 below, ordered from most to least accurate: "(1) position along a common scale, (2) positions along nonaligned scales, (3) length, direction, angle, (4) area, (5) volume, curvature, and (6) shading, color saturation" (Cleveland & McGill, 1984). The decoding of the graphical format by the reader should minimize the perceptual difficulty when they use one of the decoding tasks. The results from their experiment confirm the prediction of the accuracy level of the decoding tasks. Cleveland and McGill (1984) correctly predicted that positition judgments are more accurate than length and angle judgments.

Figure 1: Elementary perceptual tasks (Cleveland & McGill, 1984)



There could arise some perceptual issues concerning graphical communication. Humans are visual creatures, our eyes and our minds continuously work as one. These are also known as cognitive and perceptual parameters. For communication the following five aspects define graphical language: definition, construction, reading, questions and limits of the image (Bertin, 1983; Hill & Milner, 2003). In the case of graphical perception, there could arise grouping effects: (1) proximity, (2) good continuation, (3) similarity, (4) common fate, (5) good form, and (6) compatibility principle (Hill & Milner, 2003). Crucial for perception of a graphical format is detection, that a reader has the ability to detect the aspect (the ratio of height to width of a graphical plot) before any other perceptual task is possible. The accuracy of visual comparison decreases, when the distance between two graphical data increases (Hill & Milner, 2003). Thus there is trade-off between detection and distance.

Matching task complexity

Graphical communication and cognitive perception are important paramenters in graphical reporting. Graphical reporting can improve the effectiveness and efficiency of communication and data analysis (Cleveland & McGill, 1984; Hwang, 1995; Frownfelter-Lohrke, 1998). Efficient use of graphs can be more allevated by task complexity (Hill & Milner, 2003). This efficiency can be allevated by matching the question that needs to be answered with the presentation format of an information set (Bertin, 1983; Davis, 1989; Vessey, 1991; Hwang, 1995). According to Bertin (1983) communicative performance, matching, depends on three ingredients: the question that needs answering, the presentation format, and the presented information set (Davis, 1989; Hill & Milner, 2003).

Another theory on matching task complexity to the effieciency of graphical reporting is "cognitive fit", introduced by Vessey (1991). According to her "cognitive fit" of task complexity is the solution for the different impact of tabular and graphial formats on decision making and therefore "cognitive fit" will enhance decision making. "Cognitive fit" means that the "problem representation" matches the "problem solving task" and this facilitates the problem solving process which will result in better decision making. If there is a mismatch problem solvers need to transform either the problem representation or the task to derive a solution, which is less efficient and effective than a match. The most suitable presentation format is the one that requires the least amount of effort from the user (Hill & Milner, 2003), and leads to more accurate judgments (Cleveland & McGill, 1984).

Guidelines for graphical displays

As I discussed earlier there are many snags to graphical reporting. Graphical reporting can be very efficient and effective, but then the task has to match the financial information presentation format. This can be difficult since there are many graphical formats to choose from. Hence, Hill and Milner (2003) developed guidelines for graphical displays which consist of three stages.

The first stage focuses on communication, specifically: "To graph or not to graph?" (Hill & Milner, 2003). It is important to determine what the purpose of the graph is. This could influence the type of relevant data and the graphical format. Another consideration is who the audience will be. The audience might have a specific knowledge base and demands to which the graphical design should be altered. So it is important to determine what the purpose is and who the audience is and maybe, after deliberation, another reporting format (e.g. tabular format) might be better suited. A tabular might be better suited for the more complex data sets (Davis, 1989; Vessey, 1991). That is the final consideration of the first stage.

In the second stage the appropriate graphical format will be chosen. This entire stage is all about exploring what the suitable graphs are. A pitfall for many designers is that they generate common graphs. However, it is important to consider the nature of the phenomenon that is being measured and which measurement scale to use (Hill & Milner, 2003). An appropriate format for continuous data is for

example a line graph. Another measurement scale is nominal, a graphical format that is appropriate for this scale is the bar chart. The judgment expected of the reader should be included in the final choice of the graphical format. Since the reader needs to decode the graphical format, their perception should also be a factor.

After the first two stages the graphical format is clear and in this final stage it comes down to perfecting the design. The principles for good graphing consist of several criteria: good design, accuracy, simplicity, clarity, fidelity and information impact (Hill & Milner, 2003). Some important steps in this stage are to title the labels and axis, legends, color use, scaling, inclusion of a zero point and grid lines. All of these steps are important to obtain, especially, clarity from a graph. Clear graphs aid a clear understanding of the essence of the results portrayed by the graphs. Therefore it is important to keep in mind that a graphical format shows the bigger picture.

Coloring

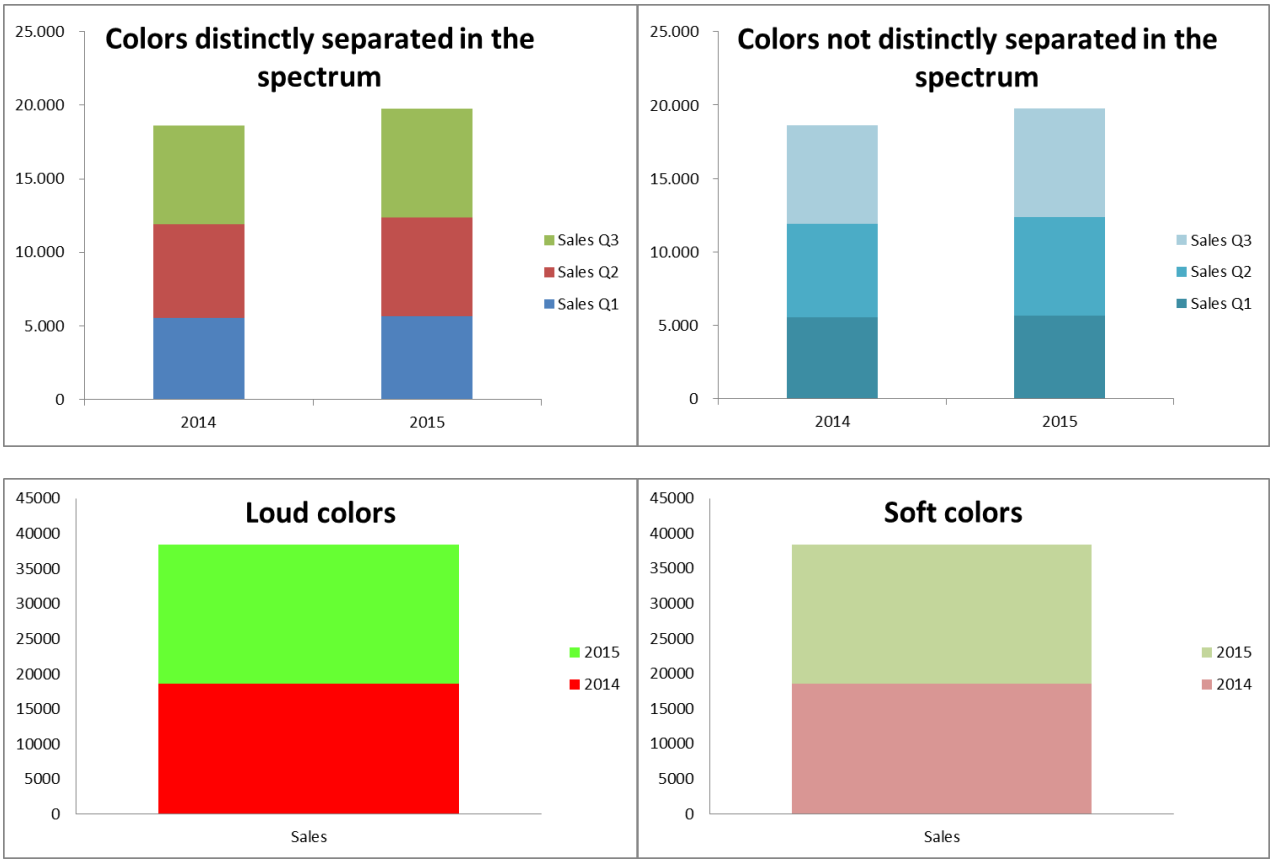
An important aspect of perfecting the design of the chosen graphical format is the use of color. Color is useful for a number of reasons: motivation, aestetics, comprehension, and recall and retention (So & Smith, 2002). Hence color can be used in a functional way (e.g. to code information) or in a practical way (e.g. to attract attention). If the color is removed from the graphical format, when the functional way of adding color is applied to visual communication, the communication will degrade. This doesn't apply to the practical way of adding color to a graphical format. If color has the function of making the graphical more appealing to readers, removing the color won't have an impact on the information content of the communication (Dooley & Harkins, 1970). Therefore an effective color code might facilitate the reader to retain important information, to organize and categorize the information and ignore redundant information.

Color coding information can be done redundant and non-redundant. The color coding is non-redundant if the information can be determined only in terms of their color (So & Smith, 2002). The color coding is redundant if the information can be determined by either the color coding or by another feature (e.g. pattern). So, the non-redundant color coding forces the reader to fully process the color, rather than corresponding data labels (So & Smith, 2002). So and Smith (2002) examine the decision performance of subjects that use non-redundant colour coding versus black and white graphics. This relation is moderated by high and low task complexity. They expect that the decision performance with high information complexity will be better with color graphics and with low information complexity the decision performance of color graphics will not be better than the decision performance black and white graphics. However their results show an opposite picture. Color coding is very effective for tasks with a low information complexity, more effective than black and white graphics. There is no performance difference for the two graphics with tasks of high information complexity, hence it could be that a graphical format is not suited for such tasks.

Another variable in the experiment of So and Smith (2002) is gender. Color is more appealing to females than males (So & Smith, 2002) and therefore they expect that the decision performance of females will be better with non-redundant color graphics. For males they expect that there won't be a difference in decision performance. The results support their expectations. The female subjects accomplish higher accuracy levels with color graphics than with the black and white versions, for men there is indeed no difference. Therefore it is possible that females and males are different in their processing of color information (So & Smith, 2002). Color is very useful to motivate and attract users and since that appeals more to females, they are more prone to use the color codes.

Since color can be applied in several ways it is important to consider which color to use. Hill and Milner (2003) explain that it is better to use colors that are distinctly separated in the spectrum, see figure 2 below. Preferred are loud colors because softer colors blend more easily and therefore do not result in good visual bias (Hill & Milner, 2003). Since colors also can communicate information, highlighting a specific feature with warm colors (e.g. red or orange) is recommended. This will appear to be in front of other information which is displayed in cooler colors (e.g. violet, green or black).

Figure 2: Examples on the use of colors



Sparklines

A graphical format that isn't previously mentioned is a little data line introduced by Tufte (2006) as sparklines. Sparklines is a new graphical format described as small, high-resolution graphics usually inserted in a full text of words, images and numbers (Tufte, 2006). Tufte (2006) uses three words to define sparklines: (data) intense, (design) simple and word-sized graphics (figure 3).

Figure 3: Application of sparklines for financial and economic data[1].

Sparklines are useful for continous data, since they present a combined pattern, a general shape and plenty of detail. Sparklines can be read as words, however a reader should read them more carefully and slowly to interpret the data. Considering, sparklines show an overall trend along with some detail they can be used for economic and financial data. Sparklines could even be incorporated in the data table for economic and financial data (e.g. exchange rates). Sparklines can portray recent change in relation to many past changes, therefore this could result in better decisions since recency bias is reduced.

Sparklines sound as a simple little line, however the design is just as important for sparklines as for other graphical formats. An important feature for the perception of the sparkline is the aspect ratio, the ratio of height to width of a graphical plot (figure 4 and 5). The aspect ratio of a sparkline is contrained by their wordlike shapes, they have on short dimension and one long, like a narrow ribbon (Tufte, 2006). The first criteria for the aspect ratio of sparklines is: use the maximal vertical space available, due to the word-shape constraint. The second criteria is the horizontal shape, the time-scale, this is stretched to meet the "lumpy criterion". This "lumpy criterion" implies that variations in slopes are best distinguished when they are around 45° (Tufte, 2006). Another feature is dequantification, which means the exclusion of labels and scaling of the general statistical displays. Since sparklines are small there is no room for these excess displays. Thus color decoding is a (contextual) method to quantify the sparklines. To design sparklines it is essential to keep the specific sparklines ideology in mind, however the general guidelines can support the design process.

[1] Page 50 in Beautiful Evidence by E. R. Tufte. In this example some color is used to help link the sparkline with the numbers.

Figure 4: Applications of the aspect ratio[2].

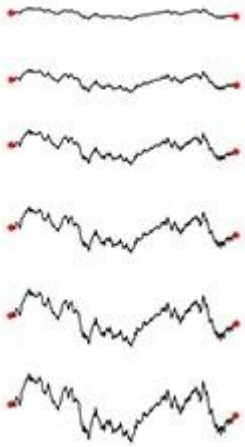
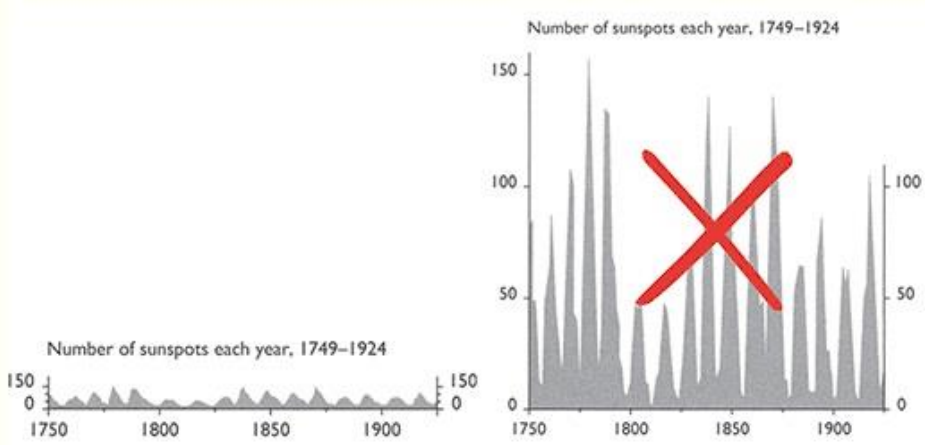


Figure 5: Application of the aspect ratio[3].



It is difficult for users to appropriately identify the slopes of the data from sparklines, since the aspect ratio is low. Althought it's difficult, it isn't impossible for users. They are able to perceive patterns in the data and recognize anomalies (Parsons & Tinkelman, 2013). Parsons and Tinkelman (2013) added semi-monthly data in graphical form (sparklines) to income statements. Their restults show that especially with pattern comparison tasks the subjects flourished when they used sparklines in addition to the tables. In addition, sparklines can provide information about the variability and timing of specific financial statement items (Masnick & Tinkelman, 2015). Specifically, material revenue and expense items. This demonstrates that sparklines can indeed be used for financial data.

[2] Page 60 in Beautiful Evidence by E. R. Tufte. In this example the x-scale (width) is held constant, while the y-scale (height) is increased by 25%. According to the lumpy criterion, the third sparkline would be a correct application.

[3] Page 60 in Beautiful Evidence by E. R. Tufte. In this example the graphics on the left meet the lumpy profile, the graphics on the right has a spiky profile.

2.2 Presentation Format

There are several papers that discuss the effect of various presentation formats that present information to users. Many research in this field was performed around 1990, however also in more recent years there is a focus for the effect of presentation formats. There is a large focus on the 'battle' between the tabular format and the graphical format. I have discussed the graphical format comprehensively, so now the focus is on the differences between these formats. Specifically, which format is better for decision making and what causes the differences between these formats.

Differences tabular and graphical format

When one looks at a graphical format and a tabular format one can immediately see the visual difference between the formats. A graphical format comes in many different shapes and designs, but they all have the distinct feature that they attract attention. Therefore, Davis (1989) concludes that the design of a graphical format has a stronger impact compared to a tabular format. The more practical use of a graphical format, however, is that they provide a holistic view, it portrays the 'big picture' and emphasizes relationships in the data (spatial) (Vessey, 1991). The tabular format, on the other hand, offers an analytical view (symbolic) and presents numerical values.

Since the formats have a different presentational feature the research first of all focuses on why the impact of the two formats is different. Around 1990 the general thinking was that task effects caused the unexpected results in the controversy, graphs versus tables. However, both task type and task complexity still offered mixed results as a mediating variable. That is when Vessey (1991) introduces cognitive fit, this means that the "problem representation" matches the "problem solving task" and this facilitates the problem-solving process which will enhance decision making. According to Vessey (1991) the graphical format presents spatial problem representations and the tabular format symbolic problem representation. The graphical format could therefore help solve spatial tasks, assessing the problem as a whole (e.g. compare two data points), which matches the spatial problem representations portrayed by a graph. The tabular format however could solve a symbolic problem, such as extracting specific data, which matches the symbolic problem representation of the tabular format. Vessey (1991) concluded that her theory does explain the majority of the results of the published studies on the graphs versus tables predicament, she used in her empirical study, and is therefore validated. Although the theory of cognitive fit is validated by the empirical study of Vessey (1991), Frownfelter-Lohrke (1998) had a different conclusion. She investigated three issues of which one was cognitive fit. Specifically she examined if decision quality will be better when the task and format match (Frownfelter-Lohrke, 1998). The results of her experiment failed to support the theory of cognitive fit.

Task complexity (and time pressure)

Even though there are mixed results on cognitive fit, task complexity is often a variable in the tabular versus graphical controversy, 'which format is more effective?'. Davis (1989), Hwang (1995), and So and Smith (2004) all use task complexity in experiment. Davis (1989) goal is to examine whether the information a decision maker wants to obtain from a report format is a task characteristic. He also wants to identify the appropriate report format through the task characteristic, levels of complexity. Davis (1989) concludes that the most relevant format of presenting information depends on the question of the decision maker. He also concludes that the performance with a tabular format will be either superior or equal to a graphical format for most questions. The performance with a graphical format will only be better if there are specific visual suggestions which benefit the answer to the question.

Hwang (1995) also includes time pressure as a variable in his experiment, because time pressure affects decision making and there wasn't any empirical evidence yet. The graphical format will lead to shorter decision times (Hwang, 1995; Masnick & Tinkelman, 2015), therefore when there is more time pressure the subjects who use the graphical format will perform better. The results of the experiment of Hwang (1995) support his hypotheses. The empirical evidence consists of three experiments, for every experiment the task becomes more difficult, so the task of experiment three is the most difficult. There is also differentiation between symbolic task (H1), spatial tasks (H2), and a combination (H3). For symbolic tasks the tabular format is expected to be more effective and it is expected that the graphical format is more effective for spatial tasks. However, time pressure could play a part. The results of Hwang (1995) show that the advantage of tables for symbolic tasks disappeared when the time pressure increased. He concludes that his research contributes to the available literature by showing that graphics are more superior than tables in supporting a decision maker under medium or high time pressure.

In this more recent study by So and Smith (2004) multivariate graphics get the incremental benefit over tabular formats. They compare the combination of graphical and tabular with a tabular format on its own. Their goal is to determine the most effective presentation format for a decision-makers performance on multivariate decision tasks of varying information complexity (So & Smith, 2004). The tabular format is familiar for users and therefore there would only be an incremental benefit when the information complexity is high. So when the information complexity is low there won't be a difference in decision performance for the combination formats and the tabular format alone. However, the results are contradictory to the results of earlier studies. The usefulness of the graphical format, over the tabular format, in complex situations is not supported (So & Smith, 2004). When decision makers are provided with a graphical format, which complements the tabular format, the decision makers will use the graphical format but this will result in lower outcomes.

Combination format

Since the graphical format and the tabular format have their unique characteristics, combining the two formats might result in an 'ultimate' presentation format. So it is a format that includes both a table and a graphical design. So and Smith (2004) already concluded that when users receive a combination format they will use the graphical format provided, but this will result in lower outcomes. However the results of other research shed a different light on the subject.

The results of Stock and Watson (1984) suggest that accountants might consider altering the current format for displaying accounting information. What they specifically examine is twofold. First, they compare the accuracy of subjects by examining the impact of a change in the reporting format on the subjects' ability to detect a change in the financial status of the firm (Stock & Watson, 1984). Second, they compare the classification accuracy of humans to a decision model. This is done by comparing the classification accuracy of (1) subjects using only the facial format and (2) subjects using both the facial format and the estimates of the decision model to (3) employing a multiple discriminant analysis model (Stock & Watson, 1984). Human judgment accuracy is determined by how and to what extent an individual detects the important properties of a judgment task and applies the judgment method. This accuracy should, in theory, improve from the use of graphical formats as substitutes for the tabular format or as an addition to the tabular format. This is supported by the results of Stock and Watson (1984), subjects who received the information in graphical format were more accurate that those who receive the tabular format.

Desanctis and Jarvenpaa (1989) conclude that there is promising support for the use of a graphical format. Just as Stock and Watson (1984), Desanctis and Jervenpaa (1989) include human judgment accuracy. In general, what they want to examine is how forecasting judgment performance is affected by different presentation formats. Specifically, they examine the accuracy of subjects in forecasting financial statement information. Even though there is promising support for the use of a graphical format, there is limited support for the dispute that the graphical and combined format provides an incremental value over the tabular format in forecasting financial statement information (Desanctis & Jarvenpaa, 1989).

Another issue Frownfelter-Lohrke (1998) investigates is the efficacy of combination formats. The expectation is that decision quality will be better in symbolic and spatial tasks when the formats, tabular and graphical, are combined instead of alone. This relationship is also presented by Lucas (1981), when the formats were combined users developed a better understanding of the task than when the formats were presented separately. The accuracy of the combination group is indeed the best, however it is not significant different from the separate formats. So there is some evidence that subjects receive the best information from the combination format in different types of tasks. Therefore users of financial statements might benefit, make better decisions, from a graphical presentation in addition to the tabular statements.

Learning curves

The dominant format in financial reporting is the tabular format. Therefore decision makers are familiar working with this format. There is evidence that decision makers might benefit from a combination, but it could also result in lower outcomes. An explanation for this contradiction might be learning curves, since decision makers aren't used to working with the graphical format there is room for improvement. Frownfelter-Lohrke (1998) examined if the decision quality of decision makers will be better for the graphical format on the last trial than on the first trial, but their results show no learning effect due to presentation format. Desanctis and Jarvenpaa (1989) did find some support for the learning curves, in the final trial a significant value is obtained indicating a performance advantage from more practice. These results don't give any clarity towards learning curves as an explanation for the (in)effectiveness of the graphical and combination format.

Nature of the task

In the dispute of which format is better another variable that plays a part is the nature of the task (task type). Both task type and task complexity offered mixed results as a mediating variable. As it turns out task complexity at a certain level mediates the differences between the two formats and also the combination format is introduced of which the results aren't perfectly aligned. The variable not yet discussed is the nature of the task. Sullivan (1988), Amer (1991), and Harvey and Bolger (1996) discuss the influence of the nature of the task on presentation format.

Sullivan (1988) concludes that the nature of a task is an essential aspect of whether a presentation format will or will not be salient, because if a presentation format is salient it will influence judgments. A salient effect occurs when individuals focus more on trivial bits of information. So managers will focus more on the tabular format alone, when presented with a graph, for other tasks than prediction tasks. Managers would be more inclined to work hard on prediction tasks than on other tasks, because of the consequences on one's credibility. Other tasks are not as consequential as prediction tasks. However, managers will still analyze all the data available to them. So the nature of the task should match with the presentation format in order to avoid decision makers to focus on trivial bits of information. Amer (1991) adds that it is important for designers of displays to take into account the nature of the decision task when designing the display of financial information (Amer, 1991). In his experiment accuracy and speed were both superior when the decision task matched the presentation design.

The results of the experiment conducted by Harvey and Bolger (1996) show a clear distinction of which format is best suited for a certain task. They examine the influence of presentation format on the accuracy of judgmental forecasts. There were two types of series: untrended and trended (positively and negatively), which the subjects had to forecast. They concluded that there is a slight advantage for the

tabular design with forecasting untrended series, because there is less inconsistency and overforecasting bias with the tabular design. For the forecasting trended series they conclude that the graphical design is better, due to less trend-damping with this design. So the nature of the task influences the choice in presentation format, which is similar to the results of Sullivan (1988) and Amer (1991).

Knowledge and experience

Something similar to the learning curve is the accounting knowledge and experience of decision makers. Experienced users of financial reporting are familiar with the tabular format and therefore less experienced users could benefit from adding the graphical format or replacing the tabular format by a graphical format. It is likely that the cognitive burden will reduce with a graphical format, therefore the performance of decision makers with a limited level of accounting knowledge should improve. Experienced decision makers, who have a high level of accounting knowledge, have an analytical focus for the domain they have a great understanding in. They will perform better with tables, given that they will look for specific details, so they will have a lower performance with the graphical format.

Cardinaels (2008) examined the cost-accounting knowledge in relation to two presentation formats in cost-based decision making. At that time there was no literature on how the decision-makers with different levels of accounting knowledge are affected by the presentation of information in different reporting formats. Also the literature is still inconclusive about the relative impact of tabular versus graphical formats. He conducted an experiment in which 55 students played the role of a firm's manager. Important to note is that his graphical format did not contain all the information displayed in the table, but the most crucial of it is covered. So in essence the presentation formats are identical in information content. From the analysis of the experiment he concludes, first, that different presentation formats influence the profitability differently. The direction of the profitability is influenced by the decision makers' level of accounting knowledge. Second, for managers there is no unique way to present information, this is dependent on their level of accounting knowledge. Third, this study has shown that knowledge is an important attribute that explains the relation between a presentation format and the resulting profit improvement.

2.3 Summary

Graphical reporting

Graphical reporting is a common financial reporting practice to use graphical formats to communicate corporate information. The graphical format has some distinct features, it portrays the 'big picture', it emphasizes relationships in the data, and it is an eye catcher. Compared to the tabular format, the graphical format has a stronger impact. Graphical communication and cognitive perception are important paramenters in graphical reporting. Graphical reporting can improve the effectiveness and efficiency of communication and data analysis (Cleveland & McGill, 1984; Hwang, 1995; Frownfelter-Lohrke, 1998). Since there are many snags to graphical reporting Hill and Milner (2003) developed guidelines for graphical displays which consists of three stages: (1) to graph or not to graph, (2) appropriate graphical format, and (3) perfecting the design. An important aspect of perfecting the design of the chosen graphical format is the use of color. Color can be used in two ways, it is either functional (e.g. to code information) or practical (e.g. to attract attention).

A graphical format that isn't previously mentioned is a little data line introduced by Tufte (2006) as sparklines. Sparklines is a graphical format that can be described in three words: (data) intense, (design) simple, and word-size graphic. An important feature for the perception of the sparkline is the aspect ratio, the ratio of height to width of a graphical plot.

Presentation format

Around 1990 the general thinking was that task effects caused the unexpected results in the controversy, graphs versus tables. However, both task type and task complexity still offered mixed results as a mediating variable. A solution was presented by Vessey (1991) which is known as "cognitive fit". Although the theory of cognitive fit is validated by the empirical study of Vessey (1991), Frownfelter-Lohrke (1998) had a different conclusion. The results of her experiment failed to support the theory of cognitive fit.

Even though there are mixed results on cognitive fit, task complexity is often a variable in the tabular versus graphical controversy, 'which format is more effective?'. Davis (1989) concludes that the most relevant format of presenting information depends on the question of the decision maker. Hwang (1995) also includes time pressure as a variable in his experiment, the results show that the advantage of tables for symbolic tasks disappeared when the time pressure increased. The results of So and Smith (2004) are contradictory to the results of earlier studies. The usefulness of the graphical format, over the tabular format, in complex situations is not supported (So & Smith, 2004).

Since the graphical format and the tabular format have their unique characteristics, combining the two formats might result in an 'ultimate' presentation format. The results of Stock and Watson (1984) suggest that accountants might consider altering the current format for displaying accounting information. Subjects who received the information in graphical format were more accurate that those

who receive the tabular format. Desanctis and Jarvenpaa (1989) conclude that there is promising support for the use of a graphical format. Frownfelter-Lohrke (1998) presents some evidence that subjects receive the best information from the combination format in different types of tasks.

Another variable is the learning curve, because decision makers aren't used to working with the graphical format there is room for improvement. The results of Frownfelter-Lohrke (1998) show no learning effect due to presentation format. Desanctis and Jarvenpaa (1989) did find some support for the learning curves. However these results do not give any clarity towards learning curves as an explanation for the (in)effectiveness of the graphical and combination format.

The next variable that plays a part is the nature of the task. Sullivan (1988) concludes that the nature of a task is an essential aspect of whether a presentation format will or will not be salient, because if a presentation format is salient it will influence judgments. In the experiment conducted by Amer (1991) accuracy and speed were both superior when the decision task matched the presentation design. The results of the experiment conducted by Harvey and Bolger (1996) are similar to the results of Sullivan (1988) and Amer (1991), the nature of the task influences the choice in presentation format.

The final variable is the accounting knowledge and experience of decision makers. Experienced users of financial reporting are familiar with the tabular format and therefore less experienced users could benefit from adding the graphical format or replacing the tabular format by a graphical format. Cardinaels (2008) concludes (1) that different presentation formats influence the profitability differently, (2) for managers there is no unique way to present information, and (3) this study has shown that knowledge is an important attribute that explains the relation between a presentation format and the resulting profit improvement.

3 Hypothesis development

In this chapter I discuss the development of my hypothesis to answer my research question. In the previous chapter I discussed the relevant prior literature, this literature is the foundation of my hypothesis development.

3.1 Hypothesis development

There have been many studies on various design options for decision making under several circumstances. The designs included a tabular format, a graphical format and a combination of the tabular and graphical format also known as a combination format. A relatively new graphical format is sparklines introduced by Tufte (2006). In contrast to other graphical formats sparklines are word-sized graphics. Therefore it is difficult for users to appropriately identify the slopes of the data which is essential for the perception of users. However sparklines are (data) intense and (design) simple, opposed to the regular graphical designs. Sparklines are supposed to add a graphical dimension to, for example, economic and financial data, instead of replacing the numerical data by a common graphical format. In general, graphical formats portray the 'big picture', emphasize relationships in the data (Mackay & Villarreal, 1987; Vessey, 1991), and attract and direct the readers' attention.

The graphical format with a sparklines design has an incremental value to the tabular format. The main difference between the tabular and the graphical format is how they present information. The tabular format offers an analytical view, while the graphical format provides a holistic view. Therefore the combination of a table with sparklines does offer an attractive format. It incorporates the holistic view and the analytical view, which will also attract the readers' attention. I expect that adding sparklines to the tabular format won't result in lower outcomes as it did when So and Smith (2004) added a common graphical format to their tabular format. Sparklines purely show the overall trend along with some data, but it is difficult to identify slopes. The subjects of Parsons and Tinkelman (2013) flourished when they used sparklines in addition to tables with pattern comparison tasks. Therefore I expect that the decision making of users will be better (more accurate) with the combination format than with just a tabular format. This makes the first hypothesis:

H1: The decision making of the users is more accurate with the combination format than with the tabular format.

A variable that affects decision making is time pressure. Time pressure is included in the research of Hwang (1995). His research shows that graphics are more superior than tables in supporting a decision maker under medium or high time pressure. Even the advantage of the tabular format for the symbolic task disappeared when the time pressure was increased. This can be explained by the holistic view the

graphical format offers. This is also why, for complex situations, So and Smith (2004) concluded that decision makers will use the graphical format but it will result in lower outcomes. However, instead of giving the subjects a time limit, their time spend on an exercise will be recorded. As the results from Frownfelter-Lohrke (1998) show, the time spent on the tasks was significant and that the group with the combination format was the most accurate. Therefore it is interesting to flip the coin and to see in addition to accuracy how much time users need in order to make a decision with both formats. I expect that users will need less time to make a decision with the combination format and still be accurate. This makes the second hypothesis.

H2: The decision quality (accuracy and timing) of the users is better with the combination format than with the tabular format.

The next variable is task complexity, which is known as a mediating variable in the tabular versus graphical format literature. The theory of cognitive fit by Vessey (1991) contributes to this variable. The question asked should match the task required to answer the question. This is similar to the conclusion of Davis (1989). Davis (1989) concludes that the most relevant format of presenting information depends on the question of the decision maker. He also concludes that the performance with a tabular format will be either superior or equal to a graphical format for most questions. In contrast to Davis (1989) I examine a sparkline design in addition to a tabular format. So I expect that the combination format will be superior for most questions than just a tabular format. So and Smith (2004) researched this relation with a common graphical format, but by adding sparklines to the tabular format I expect that their lower outcomes won't happen. They differentiated between low complexity and high complexity. Hwang (1995) on the other hand differentiates between simple, medium-complex and highly complex tasks. Hwang (1995) also included time pressure as a variable. He differentiates between low, medium and high time pressure. The results of Hwang (1995) show that the graphical format is superior in the medium-complex and highly-complex task to the tabular format under medium and high time pressure. Contrary to Hwang (1995) I examine task complexity without the variable time pressure. Hwang (1995) used common graphical designs, such as bar and line charts, while I examine a sparkline design in addition to a tabular format.

As I examine task complexity without time pressure and a different graphical format, following the results of these papers, I expect minimal superiority of the combination format over the tabular format with questions of low task complexity. On the other hand I do expect that the combination format is superior with questions of high task complexity. This makes the third and fourth hypothesis.

H3: The decision making of the users with low task complexity is the same for the combination and tabular format.

H4: The decision making of the users with high task complexity is more accurate with the combination format than with the tabular format.

The distinct feature of the graphical format is that they portray the 'big picture' and emphasize relationships in the data (Mackay & Villarreal, 1987; Vessey, 1991). Graphical formats are also eye catchers, they attract and direct the readers' attention. The design has a stronger impact compared to a tabular format (Davis, 1989). With the combination format the sparklines are incorporated in the tabular format. Sparklines portray recent performance in relation to past performance, therefore this could result in better decision since recency bias is reduced. With all these aspects of the graphical format in general and specifically the possible added value to the tabular format of sparklines, I expect that decision makers can memorize this information better than when provided with solely a tabular format. It is interesting to see how accurate decision makers are when they have to memorize the information provided in either the tabular format or the combination format. As decision makers in general memorize the performance of a company to make decisions and forecasts. Therefore, I expect that the decision making of users when memorizing the information provided is more accurate with the combination format than with the tabular format. This makes the fifth and final hypothesis.

H5: The decision making of the users through memorization is more accurate with the combination format than with the tabular format.

3.2 Summary

In this chapter I have developed my hypothesis based on the literature I discussed in chapter two. There have been many studies on various design options for decision making under several circumstances. A relatively new graphical format is sparklines introduced by Tufte (2006). Sparklines are supposed to add a graphical dimension to economic and financial data, instead of replacing the numerical data by a common graphical format. The combination of a tabular format with a sparklines design offers an attractive format. It incorporates the holistic view and an analytical view, which will also attract the readers' attention. In total I have formulated five hypotheses which in additional to accuracy include time measurement, task complexity and time pressure.

H1: The decision making of the users is more accurate with the combination format than with the tabular format.

H2: The decision quality (accuracy and timing) of the users is better with the combination format than with the tabular format.

H3: The decision making of the users with low task complexity is the same for the combination and tabular format.

H4: The decision making of the users with high task complexity is more accurate with the combination format than with the tabular format.

H5: The decision making of the users through memorization is more accurate with the combination format than with the tabular format.

4 Research design

In this chapter I discuss the research design I implemented to answer the hypotheses. In the previous chapter I have developed my hypotheses which are the basis for my research design.

4.1 Experimental design

The conducted experiment is based upon a between subjects design and a within subjects design. To answer the hypotheses the between subjects design is applied. All subjects do receive the successive treatments. However, there is simultaneous variation of the treatments (see Appendix 9.1 figure 1). The within subjects design is not tested to answer the hypotheses. This is due to a possible case order effect, since the first two cases aren't switched. To clarify the treatments I will go into the experiment itself and how I set up the experiment.

4.1.1 The experiment

The experiment consists of two parts in which four companies[4] are discussed. For every company an income statement of the third quarter (Q3) of 2015 is provided for the quarterly earnings announcement. In this announcement the numbers of Q1 and Q2 are also included as are the comparative numbers Q1 till Q3 of 2014. The format of the income statement for these companies is either the combination format or the tabular format. For company A and B, which are discussed in part 1, an additional income statement is included, specifically the income statement of 2014 with the numbers for Q1 till Q4.

The first part of the experiment consists of two treatments. Treatment 1 starts with company A which provides the income statement in the combination format. The subject has to answer four questions based upon the provided income statement. The treatment then continues with company B who provides an income statement in the tabular format. Again the subject has to answer four questions based upon the provided income statement. These are the same four questions for company A as for company B. The difficulty of the questions increase, since I differentiate in task complexity in order to answer hypotheses 3 and 4. In paragraph 4.1.3 I will go further into the low and high task complexity. The difference between treatment 1 and treatment 2 is that company A now provides an income statement in the tabular format and company B provides an income statement in the combination format (see Appendix 9.1 figure 4, 5, 6 and 7).

The second part of the experiment consists of four treatments. Company C and D have provided their income statement in the combination format and the tabular format. The first treatment shows the

[4] The companies (A to D) are loosely based upon existing companies, however those income statements have been simplified and altered (the amounts have been doubled or halved), so subjects won't be able recognize a company or trace it back to the original company.

income statement of Company C in the combination format and next to it the income statement of Company D in the tabular format. These income statements are shown to the subject for a limited amount of time, 120 seconds. On the next two pages there are four questions based upon the provided income statements, on the first page the questions for company C and on the second page the same questions for company D. The difference between the four treatments is shown in Appendix 9.1 figure 8.

4.1.2 The experimental task

Before the subjects enter into the experiment they receive a brief introduction (Appendix 9.1 figure 2). For the experiment the subjects will adopt the role of investor for the four companies. They take upon this role, because I am interested in seeing how well decision makers perform with different presentation formats. The investor is in general a stake holder in a companies' performance. A company shows its performance through their financial statement, which includes the income statement. Through quarterly earnings announcements the company informs its stakeholders quarterly instead of only annually by issuing their (annual) financial statement. Investors base their investment decisions for a start on the performance of the company through their earnings announcement. That is why the subjects will adopt the role of investor in this experiment.

For each firm the subjects answer questions about the performance of the firm. To interpret the performance of the firm the subjects use the provided income statements. For the first treatment of the experiment this are the four questions (see Appendix 9.1 figure 9):

1. Did the income from operations increase or decrease in 2015?
2. Compare the sales of 2015 with the sales of 2014. Which of the following best describes the patterns of these two items?
3. What has been the performance of Company A in the first nine months of 2015 relative to the first nine months of 2014?
4. Based on the provided income statement and the additional information provided above, what are your predictions for the fourth quarter of 2015? (for sales, income from operations and net income)

These questions require subjects to compare numbers, compare trends and predict trends. For the first question the subject only has to interpret the number of Q1 and Q3 of 2015. Questions 2 and 3 are more difficult to answer, since the subjects have to compare 2015 to 2014 and interpret the trends and numbers. For question 4 also predicting numbers is added to the task.

The contrast with treatment 1 is that for treatment 2 the subjects have to interpret the financial position of company C and D without knowing precisely on what to focus, since they have to answer the questions after they have seen the income statements. They have 120 seconds to memorize the income statements. Once they leave the page, even before the 120 seconds have passed, they can't return to the income statements. The subjects do know that they have to make a decision as an investor and

therefore keep in mind the performance of the firm in 2014 and 2015. The questions the subjects have to answer for treatment 2 are (see Appendix 9.1 figure 10):

1. Did the sales increase or decrease in 2015
2. Did the income from operations increase or decrease in 2015?
3. Did the net income increase or decrease in 2015?
4. What has been the performance of Company C in the first nine months of 2015 relative to the first nine months of 2014?

The focus is on remembering the financial position of the companies and answering questions for which the subjects have to compare numbers and compare trends focused on 2015 and only once they have to make a comparison with 2014 under some time pressure.

4.1.3 Task complexity

As I mentioned in paragraph 4.1.2 the difficulty of the questions increase. In this paragraph I explain how I use increasing task complexity to manipulate the experiment, as a mediating variable. As Hill & Milner (2003) explained efficient use of graphs can be more alleviated by task complexity. This efficiency can be allevated by matching the question that needs to be answered with the presentation format of an information set (Bertin, 1983; Davis, 1989; Vessey, 1991; Hwang, 1995). I developed questions that can be answered by using both formats, but what I want to test is if the decision making of users, in the experiment investors, is affected by the competing design for the income statement of the quarterly earnings announcement. All questions are closed, therefore the answers are measurable and comparable. For every type of question I will explain what steps the subject has to take in order to answer the question correctly, this will make clear why a question is of higher complexity.

Type 1[5]: Did the income from operations increase or decrease in 2015?

In order to answer this question the subject has to check the income statement line income from operations. With the combination format they can see in a blink of an eye if the income from operations increases or decreases in 2015 and with the tabular format the subject has to compare the number of Q1 with Q3. This is classified as a question of lower complexity in this experiment, since the subject only has to focus on one line item; on two numbers, or one sparkline.

Type 2[6]: Compare the sales of 2015 with the sales of 2014. Which of the following best describes the patterns of these two items?

This question isn't necessarily based on the exact numbers, but on the pattern comparison. The subjects need to compare the pattern of the sales, one line of the income statement, of Q1 till Q3 of 2014 with Q1 till Q3 of 2015. This can be easily done with the combination format by comparing the sparklines, however with the tabular format the subject has to compare the numbers of all six quarters. Therefore

[5] This type applies to question 1 of treatment 1 and question 1 till 3 of treatment 2.

[6] This type applies to question 2 of treatment 1.

this question is a different type than type 1, since it focuses on the pattern, and can therefore be classified as a question of low/medium complexity in this experiment. The subject only has to focus on one line item, but on six numbers or two sparklines.

Type 3[7]: What has been the performance of Company A in the first nine months of 2015 relative to the first nine months of 2014?

For this question some judgment is required since the subject has to determine what the performance of a company entails. In general the performance is measured by the net income/earnings and therefore the subject should base their answer on this line item of the income statement. The following steps are the same as type 2 the subject should compare the patterns. Since this question requires additional judgment in comparison with type 2, this question can therefore be classified as a question of medium complexity in this experiment.

Type 4[8]: Based on the provided income statement and the additional information provided above, what are your predictions for the fourth quarter of 2015? (For sales, income from operations and net income)

In contrast to the other types this question needs some additional information. Therefore an additional income statement is provided. This is the income statement of 2014 with the numbers for Q1 till Q4. With the additional information the subject need to use their judgment to predict the fourth quarter of 2015. The subjects not only have to make a prediction for one line item, but for three line items. In sum the subjects must focus on three line items, compare the numbers/sparklines of 2015 to those of 2014 for the first three quarters and then make a prediction based upon the pattern of 2014 by using their judgment. Since these questions require the most steps and judgment in comparison to the other types, this question is classified as a question of high complexity in this experiment.

4.1.4 Memorization

In this paragraph I explain how I use memorization to manipulate the experiment, as a mediating variable. As I explained in paragraph 4.1.2. in treatment 2 a time element is added to the experiment. This time element is that the subjects have 120 seconds to memorize the income statements. The subjects are shown the following message: “As an investor of both Company C and D you just received their earnings announcement of the third quarter of 2015. To make your analysis of their performance you use the provided income statements, but you only get a limited amount of time to study the income statements.” Dooley and Harkins gave their subjects first 150 seconds and then another 60 seconds. However, their subjects weren’t made aware that they had to memorize the graphs. In a trial run I determined that 120 seconds was enough to memorize both graphs to subsequently answer the corresponding questions.

[7] This type applies to question 3 of treatment 1 and question 4 of treatment 2.

[8] This type applies to question 4 of treatment 1.

4.2 Experimental setting

The experiment is setup online and is therefore provided to the subjects through an internet link. As I mentioned in paragraph 4.1.2 before the subjects take part in the experiment they are briefed (Appendix 9.1 figure 2) that they should complete this experiment on a computer or tablet, since the second treatment isn't correctly shown on a mobile device. The subjects that have taken part in the experiment did it in their own time. In order to track their time spent on the treatments and the experiment itself I added a timing block to the experiment. So their time spent on a single page is timed as is their total time spent on the experiment. The experimental conditions were assigned randomly.

The subjects comes across the following sections when they take part in the experiment (Appendix 9.1): (1) introduction and briefing, (2) background questionnaire, (3) treatment 1, (4) treatment 2, and (5) debriefing questionnaire.

4.3 Dependent variables

There are two dependent variables; accuracy, timing. Accuracy of a subject is measured by their accuracy in answering the questions. So this entails amount of right and wrong answers. Timing represents the time a subject spent on the treatment and the total experiment which is measured in seconds. For the Libby Boxes check the Appendix 9.2.

4.4 Subjects

The subjects are university students in their masters. Specifically, the subjects are from the master Accounting, Auditing & Control. In total forty students took part in the experiment, seventeen females and twenty-three males. The average subject is male, 23 years old and Dutch, see for the distributions Appendix 9.3 figure 13 and 14. Of those forty students half of them had an internship at an accounting firm, see Appendix 9.3 figure 15.

5 Results

In this chapter I discuss the results from the statistical tests I performed for the hypotheses. The basis for the statistical tests is the research design which I described in chapter 4. To answer the hypotheses I use the first part of the experiment for the first 4 hypotheses and the second part for the fifth hypothesis. The experiment is based on a between-subjects design, this is applied to every hypothesis.

5.1 Hypothesis 1

The first hypothesis tests the accuracy of the decision makers with the combination format and with the tabular format. The expectation is that the decision making of the users is more accurate with the combination format than with the tabular format. The accuracy is measured by the amount of questions the subject answered correctly. The dependent variable, accuracy, is measured on a continuous scale (0 – 6). The observations are independently observed. The data meets the assumptions for the independent samples t-test which I performed.

I ran an independent samples t-test for the two formats to determine if the subjects that received the combination format were statistically more accurate than the subjects that received the tabular format. The result[9] of the t-test is insignificant, $t(78) = -1.521$, $\rho = 0.132$, and therefore I can't accept hypothesis 1. The group statistics show that the subjects that received the tabular format were on average, not statistically, more accurate than the subjects that received the combination format (4.425 over 4.000 out of 6).

Independent T-Test

Group Statistics

	N	Mean	Std. Deviation
Result Case 1A & 2A	40	4.000	1.220
Result Case 1B & 2B	40	4.425	1.279

Independent Samples Test

	Mean Difference	t	Sig. (2-tailed)
Equal variances assumed	-0.425	-1.521	0.132
Equal variances not assumed	-0.425	-1.521	0.132

Hartley test for equal variance: F = 1.100, Sig. = 0.3828

[9] I assume, based on the Harley test for equal variance, equal variance.

5.2 Hypothesis 2

The second hypothesis tests the accuracy and the timing of the users with the combination format and with the tabular format. The expectation is that users will need less time to make a decision with the combination format and still be accurate. The accuracy is measured by the amount of questions the subject answered correctly. The timing is measured in seconds, the total time the subjects spend time on the page answering the questions. The appropriate test for the between-subjects design and these variables is the One-Way MANOVA. This test requires at least two dependent variables, accuracy and time. The dependent variable, accuracy, is measured on a continuous scale (0 – 6). The other dependent variable, time, is also measured on a continuous scale (in seconds). The observations are independently observed. The data meets the assumptions for the One-Way MANOVA.

The result of the multivariate test is $F(2, 77) = 1.185$, $\rho = 0.311$; Wilk's $\Lambda = 0.970$. The result is insignificant because $p > 0.05$, therefore I can't accept hypothesis 2. From the descriptive statistics I derive that on average, not statistically, subjects spend more time with the combination format and were less accurate in comparison with the tabular format (4.425 over 4.000 out of 6).

One-Way MANOVA

Descriptive Statistics

	N	Mean	Std. Deviation
Timing Case 1A & 2A	40	180.1130	158.91326
Timing Case 1B & 2B	40	172.5594	118.59137
Total	**80**	**176.3362**	**139.37097**
Result Case 1A & 2A	40	4.0000	1.21950
Result Case 1B & 2B	40	4.4250	1.27877
Total	**80**	**4.2125**	**1.25983**

Multivariate Tests[a]

		Value	F	Sig.
Intercept	Wilks' Labda	0.071	500.306[b]	0.000
Case 1A & 2A	Wilks' Labda	0.970	1.185[b]	0.311

a. Design: Intercept + Case 1A & 2A

b. Exact statistic

5.3 Hypothesis 3

The mediating variable task complexity is included to test the third and fourth hypothesis. Task complexity consists of two parts: low and high task complexity. The third hypothesis tests the accuracy of the decision makers with low task complexity for the combination and the tabular format. The expectation is that the decision making of the users is the same with the combination format and the tabular format. The accuracy is measured by the amount of questions of low task complexity the subjects answered correctly. The test I perform is the same test as I performed for hypothesis 1, the independent samples t-test. The difference is that for hypothesis 3 I only compare the answers of the questions of low task complexity. The dependent variable, accuracy, is measured on a continuous scale (0 – 2). The observations are independently observed. The data meets the assumptions for the independent samples t-test which I performed.

The result[10] of the independent samples t-test is insignificant, $t(78) = 0.542$, $\rho = 0.598$. This entails that there isn't a significant difference in the accuracy of the subjects with the two formats and therefore the hypothesis isn't rejected. The group statistics show that on average, not statistically, the subjects that received the combination format were more accurate than the subjects that received the tabular format (1.525 over 1.450 out of 2).

Independent T-Test (Low)

Group Statistics

	N	Mean	Std. Deviation
Result Case 1A & 2A Low	40	1.525	0.599
Result Case 1B & 2B Low	40	1.450	0.639

Independent Samples Test

	Mean Difference	t	Sig. (2-tailed)
Equal variances assumed	0.075	0.542	0.598
Equal variances not assumed	0.075	0.542	0.598

Hartley test for equal variance: F = 1.138, Sig. = 0.3426

[10] I assume, based on the Harley test for equal variance, equal variance.

5.4 Hypothesis 4

The fourth hypothesis tests the accuracy of the decision makers with high task complexity for the combination and the tabular format. The expectation is that the decision making of the users is more accurate with the combination format than with the tabular format when answering questions of high task complexity. The accuracy is measured by the amount of questions of high task complexity the subjects answered correctly. The dependent variable, accuracy, is measured on a continuous scale (0 – 4). The observations are independently observed. The data meets the assumptions for the independent samples t-test which I performed.

The result[11] of the independent samples t-test is significant, $t(78) = -2.195$, $\rho = 0.031$. This is however not in line with the expectation, because on average the accuracy with the tabular format is higher than with the combination format when answering question of high task complexity. Therefore hypothesis 4 is not accepted. The group statistics show that on average, not statistically, the subjects that received the tabular format were more accurate than the subjects that received the combination format (2.975 over 2.475 out of 4).

Independent T-Test (High)

Group Statistics

	N	Mean	Std. Deviation
Result Case 1A & 2A High	40	2.475	1.109
Result Case 1B & 2B High	40	2.975	0.920

Independent Samples Test

	Mean Difference	t	Sig. (2-tailed)
Equal variances assumed	-0.500	-2.195	0.031
Equal variances not assumed	-0.500	-2.195	0.031

Hartley test for equal variance: F = 1.455, Sig. = 0.1200

[11] I assume, based on the Harley test for equal variance, equal variance.

5.5 Hypothesis 5

The fifth hypothesis tests the accuracy of the decision makers when memorizing the information provided in the combination and the tabular format. The expectation is that the decision making of the users is more accurate with the combination format that with the tabular format. The accuracy is measured by the amount of questions the subjects answered correctly. The dependent variable, accuracy, is measured on a continuous scale (0 – 4). The observations are independently observed. The data meets the assumptions for the independent samples t-test which I performed.

The result[12] of the independent samples t-test is insignificant, $t(78) = 0.214$, $\rho = 0.831$. This entails that there isn't a significant difference in the accuracy of the subjects with the two formats and therefore the hypothesis is not accepted. The group statistics show that on average, not statistically, the subjects were more accurate with the combination format than with the tabular format (3.35 over 3.30 out of 4).

Group Statistics

	N	Mean	Std. Deviation
Combination format	40	3.35	1.001
Tabular format	40	3.30	1.091

Independent Samples Test

	Mean Difference	t	Sig. (2-tailed)
Equal variances assumed	0.050	0.214	0.831
Equal variances not assumed	0.050	0.214	0.831

Hartley test for equal variance: F = 1.188, Sig. = 0.2943

[12] I assume, based on the Harley test for equal variance, equal variance.

5.6 Summary

The results of the experiment don't support the formulated hypotheses, except for hypothesis 3. The results of the tests are included in the table below.

The expectation of hypothesis 1 is that the decision making of the users is more accurate with the combination format than with the tabular format. The result of the independent samples t-test is insignificant, ρ = 0,132, therefore hypothesis 1 can't be accepted.

The expectation of hypothesis 2 is that the decision quality (accuracy and timing) of the users is better with the combination format than with the tabular format. The result of the one-way MANOVA is insignificant, ρ = 0.311, therefore hypothesis 2 can't be accepted.

For hypotheses 3 and 4 a mediating variable, task complexity, is included. The expectation of hypothesis 3 is that the decision making of the users with low task complexity is the same for the combination and tabular format. The result of the independent samples t-test is insignificant, ρ = 0,598, therefore hypothesis 3 is accepted.

The expectation for hypothesis 4 is that the decision making of users with high task complexity is more accurate with the combination format than with the tabular format. The result of the independent samples t-test is significant, ρ = 0.031. However the result is not in line with the expectation because on average the accuracy with the tabular format is higher than with the combination format when answering question of high task complexity. Therefore hypothesis 4 is not accepted.

For hypothesis 5 a mediating variable, memorization, is included. The expectation of hypothesis 5 is that the decision making of users when memorizing the information provided is more accurate with the combination format than with the tabular format. The result of the independent samples t-test is insignificant, ρ = 0.831, therefore hypothesis 5 is not accepted.

Summary:

	Test	Sig.	Conclusion
Hypothesis 1	Independent Samples T-Test	0.132	Rejected
Hypothesis 2	One-Way MANOVA	0.311	Rejected
Hypothesis 3	Independent Samples T-Test	0.598	Accepted
Hypothesis 4	Independent Samples T-Test	0.031	Rejected
Hypothesis 5	Independent Samples T-Test	0.831	Rejected

6 Additional tests

In this chapter I discuss the result of the additional tests I performed. First, I discuss the results of the statistical tests, between-subjects design, to answer the hypotheses in which 4 subjects are excluded. Second, I discuss the theory behind the second part of the experiment. I also discuss the results of the independent sample t-test in order to test the theory of the second part of the experiment. Finally, I discuss the two debriefing questions of the subjects in relation to their overall performance.

6.1 Elimination of responses

In the previous chapter I determined that the outcome of the performed statistical tests don't support the formulated hypotheses, except for hypothesis 3. There are several possibilities why the results don't support my expectations. In this chapter I discuss one of the possibilities, for the other ones I refer to chapter 7. A possibility is that subjects who took part in the experiment only did it as a favor for me, but didn't took the appropriate amount of time and effort to complete it. This can result in unnecessary wrong answers which can alter the outcome of the statistical tests. To eliminate this effect I checked the normality distribution of the case results and the time spend on a case. For the distributions see Appendix 9.4, figures 16 till 22.

In order to eliminate responses I created a pivot table which included every subject and how they performed as a decision maker, which I base on their decision quality (time + accuracy). In order to determine which subject performed worse I check the normality distributions for outliers. Outliers can either be winsorized or removed. However, winsorizing isn't possible due to the nature of this study. Therefore I eliminate the divergent subjects. In the tables I highlighted the outliers. I only highlighted the outliers on the left side of the normality distribution. These are the bad results and short time spend on the case. If a subject is an outlier for both result and time on the same case, it is eliminated. The responses I eliminated I have highlighted red. For the overview check Appendix 9.4, figure 23. In addition I included the new distributions after the subjects are eliminated, see Appendix 9.5.

Re-performance of the statistical tests.

Without the four underperforming subjects I re-performed the statistical tests I performed in order to answer the hypotheses. So for hypotheses 1, 3, and 4 I performed the independent samples t-test, and for hypotheses 2 the one-way MANOVA. I will discuss the results for every hypothesis individually.

First, I discuss the results of hypothesis 1 in relation with the test performed in chapter 5.1. The group statistics increased for both formats. However, the group statistics still show that the subjects that received the tabular format were on average, not statistically, more accurate than the subjects that received the combination format (4.500 over 4.250 out of 6). The result of the independent samples t-

test hasn't changed. The result of the t-test is insignificant, $t(70) = -0.926$, $\rho = 0.357$, therefore I can't accept hypothesis 1.

Second, for hypothesis 2 I derive, from the group statistics, that on average (not statistically) subjects spend more time with the combination format and were less accurate in comparison with the tabular format (4.500 over 4.250 out of 6). This has not altered in comparison with chapter 5.2. The result of the multivariate test is $F(2, 69) = 0.471$, $\rho = 0.627$; Wilk's $\Lambda = 0.987$. The result is insignificant because $p > 0.05$, therefore I can't accept hypothesis 2.

Third, the result for hypothesis 3 isn´t different from the result discussed in chapter 5.3. The group statistics show that on average, not statistically, the subjects that received the combination format were more accurate than the subjects that received the tabular format (1.556 over 1.472 out of 2). The result of the independent samples t-test is insignificant, $t(70) = 0.582$, $\rho = 0.562$. This entails that there isn't a significant difference in the accuracy of the subjects with the two formats and therefore the hypothesis isn't rejected.

Fourth, in comparison with 5.4 the result of hypothesis 4 is altered. The group statistics show that on average, not statistically, the subjects that received the tabular format were more accurate than the subjects that received the combination format (3.028 over 2.694 out of 4). The result of the independent sample t-test is insignificant, $t(70) = -1.573$, $\rho = 0.120$, therefore I can't accept hypothesis 4.

Fifth, and final, the result for hypothesis 5 isn't different form the result discussed in chapter 5.5. The group statistics show that on average, not statistically, the subjects were more accurate with the combination format than with the tabular format (3.53 over 3.42 out of 4). The result of the independent samples t-test is insignificant, $t(64.94) = 0.489$, $\rho = 0.626$. This entails that there isn't a significant difference in the accuracy of the subjects with the two formats and therefore the hypothesis is not accepted.

Group Statistics

	N	Mean	Std. Deviation
Result Case 1A & 2A	36	4.250	0.996
Result Case 1B & 2B	36	4.500	1.276
Timing Case 1A & 2A	36	197.7807	157.72469
Timing Case 1B & 2B	36	183.2476	117.97172
Result Case 1A & 2A Low	36	1.556	0.558
Result Case 1B & 2B Low	36	1.472	0.654
Result Case 1A & 2A High	36	2.694	0.920
Result Case 1B & 2B High	36	3.028	0.878
Result Case 3&4 - Combination format	36	3.530	0.810
Result Case 3&4 - Tabular format	36	3.420	1.079

Independent Samples Test[13]

	Mean Difference	t	Sig. (2-tailed)
Hypothesis 1	-0.250	-0.926	0.357
Hypothesis 3	0.083	0.582	0.562
Hypothesis 4	-0.333	-1.573	0.120
Hypothesis 5	0.110	0.489	0.626

Multivariate Tests[a]

		Value	F	Sig.
Intercept	Wilks' Labda	0.050	652.872[b]	0.000
Case 1A & 2A	Wilks' Labda	0.987	0.471[b]	0.627

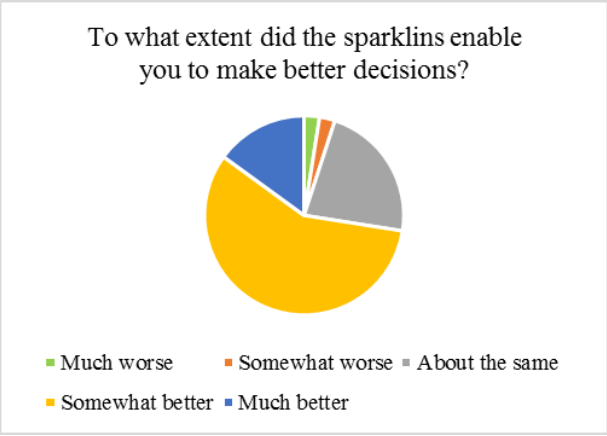
a. Design: Intercept + Case 1A & 2A
b. Exact statistic

In relation to the prior tests I performed, the result of the re-performed statistical tests (without the four underperforming subjects) isn't different than the result of chapter 5.6. The results don't support the formulated hypotheses, except for hypotheses 3 and 5. The only difference is that hypothesis 4 is now insignificant. However, the conclusion hasn't changed.

6.2 Debriefing

As a conclusion to the experiment the subjects received two debriefing questions, included in Appendix 9.1, figure 11. The goal of the questions is to get an insight of how the subjects experienced the sparklines. In figure 6 and 7 the distribution of the answers to the questions is portrayed.

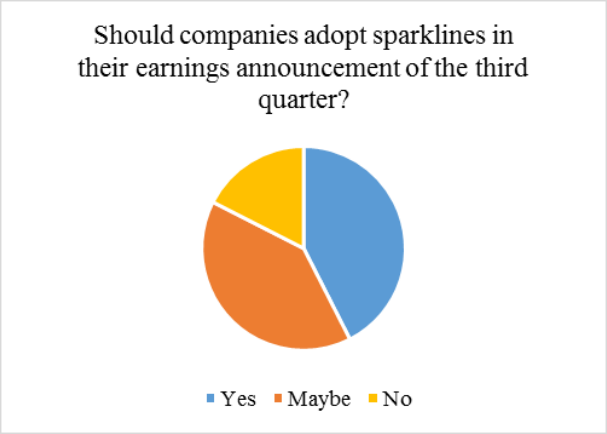
Figure 6: Answer distribution to the first question.



[13] I assume, based on the Harley test for equal variance, equal variance for hypotheses 1, 3 and 4.
- H1: Hartley test for equal variance: F = 1.640, Sig. = 0.0712.
- H3: Hartley test for equal variance: F = 1.375, Sig. = 0.1719.
- H4: Hartley test for equal variance: F = 1.099, Sig. = 0.3894.

I assume, based on the Harley test for equal variance, no equal variance for hypothesis 5.
- H5: Hartley test for equal variance: F = 1.774, Sig. = 0.0448.

Figure 7: Answer distribution to the second question



When the subjects were asked to what extent sparklines enabled them to make better decisions 72.5 percent answered positively. Even 15 percent of the subjects was even much better facilitated with sparklines. Only 5 percent of the subjects said that they made worse decisions when they were presented sparklines. These positive answers are also reflected in the second question. The subjects are mostly divided between the answers 'yes' (42.5%) and 'maybe' (40%). Only 17.5 percent of the subjects answered 'no' to the question if companies should adopt sparklines in their earnings announcement of the third quarter.

However, this positivity doesn't entail that the subjects also performed better (statistically) with sparklines as is noted in chapter 5. Therefore, I first determine with which format the subjects are more accurate, this is shown in figure 8. Secondly, I determine whether the subjects had the right conclusion in the debrief about their performance, this is shown in figure 9.

Figure 8: Accuracy subjects in the first treatment.

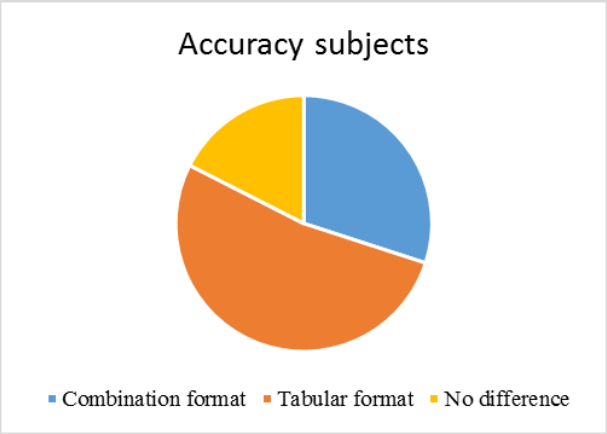
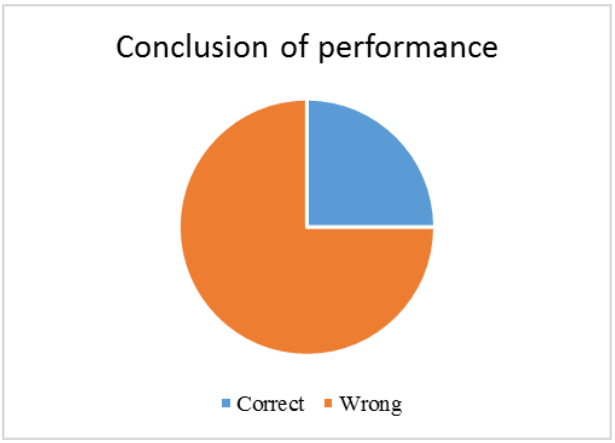


Figure 9: Conclusion on the performance compared to the answer in the debrief.



In general, my observation is that 52.5 percent of the subjects was more accurate when they were presented with the tabular format than with the combination format. There were also subjects, 17.5 percent that performed as well with the tabular format as with the combination format. As figure 9 portrays only 25 percent of the subjects had the right conlusion in the debrief. There are subjects that performed better with the combination format and concluded that sparklines didn't enable them to make better decisions and the other way around. From this observation I conclude that the positive conclusion of the subjects towards sparklines isn't backed up by their performence in the experiment.

7 Discussion

An experiment with low and high task complexity was conducted to test whether a competing design for the income statement would affect the decision making of users of the financial statements. I expected that the accuracy of users would be better especially with high task complexity. I also expected that the decision quality of users would be better with the competing design instead of the tabular format. However, the results do not support my expectation except for questions of low task complexity. Even after I eliminated four subjects the results did not alter. Therefore, I can't say that the competing design positively affects the decision making of users of the financial statement.

Even though I did not find a statistical confirmation that the combination format positively affects the decision making of the users of the financial statement, I do make several contributions with this research. I found statistical evidence that there is no difference between the two formats under low task complexity. Before the additional analysis I found statistical evidence that under high task complexity the accuracy is different for the two formats. I found that on average the accuracy with the tabular format is higher than with the combination format when answering question of high task complexity. However, with the additional analysis the result of accuracy under high task complexity is insignificant, there is no difference between the two formats under high task complexity.

Additionally I observed that the subjects are positive about sparklines. This is interesting for policy makers and companies since users of financial statements are positive about sparklines. I also doesn't have a negative impact on their decision making. Therefore, implementing the combination format won't result in less accurate decision making. However, further research regarding sparklines is recommended to determine the efficiency of the combination format and the implementation in the financial statement.

That the outcome of my research question isn't in line with my expectations can be due to several factors. The first factor is that I am also following the master Accounting, Auditing & Control as do most of the subject. Therefore, some of the subjects who took part in the experiment are friends of mine. It is possible that they took part in the experiment as a favor to me. This can entail, as I explained in chapter 6, that they didn't took the appropriate amount of time and effort to complete the experiment successfully. This can result in unnecessary wrong answers which can alter the outcome of the statistical tests. For future experiments I would advise to give an incentive to participants, as researchers did in previous studies. For example extra credit or a that subjects can win a gift certificate. In addition I would advise to expand the sample size. I had a limited sample size of 40 subjects. This can be enough. However, the results will be stronger when the sample size is expended.

In the second place the experimental cases weren't cohesive. The line items in the tables in case A and B were different. In case A I used the terminology Sales, Income from operations and Net income which aligned with the questions. However, in case B I used different terminology than in case A to

express the same line items. The terminology I used is Revenue, Operating income and Net earnings. This difference in terminology from the questions to be answered could have confused subjects. This confusion could result in lower outcomes. I would therefore recommend future researchers to either thoroughly test that terminology isn't a significant influence. Or to use the same terminology in the different cases.

Thirdly, I initially based my experiment on a between- and a within subjects design. However, the within subjects design is not tested to answer the hypotheses. This is due to a possible case order effect, since the first two cases aren't switched. Subjects always received case A first and then case B. Additionally, they either received case A with sparklines and then case B without sparklines or case A without sparklines and case B with sparklines. If I would have focused on only one research design, a between subjects design, my experiment would be more complete. Therefore, I would recommend for future research to explore the combination format with either a between or a within subjects design.

For the within subjects design I have another recommendation. The dominant format in financial reporting is the tabular format. As prior research showed there is evidence that decision makers might benefit from a combination, but it could also result in lower outcomes. An explanation for this contradiction might be learning curves, since decision makers aren't used to working with the graphical format there is room for improvement. Frownfelter-Lohrke (1998) and Desanctis and Jarvenpaa (1989) didn't find supporting evidence for learning curves. However, they didn't work with the sparkline design. I noted that the results in the second case were on average (not statistically) higher for both formats and the subjects were on average (not statistically) faster for both formats. Thus, I recommend for future research to explore the learning effect of the accuracy of users of the financial statement, with the combination format in contrast to the tabular format in a within subjects design,

As a final remark I focused mainly on the accuracy of the users with low task complexity. As I expected and established there was no difference in accuracy between the two formats. This is in line with prior literature that used different graphical formats in their research. As prior research showed there is some evidence that higher task complexity results in better decision making with the combination format. I would recommend future research to explore task complexity as a mediating variable further combined with time pressure. Specifically, to re-perform the study of Hwang (1995) but then with sparklines for the combined format. This would be interesting for policy makers and companies since users of financial statements are positive about sparklines.

8 Bibliography

Amer, T. (1991). An experimental investigation of multi-cue financial information display and decision-making. *Journal of Information Systems, 5*(2), 18-34.

Bertin, J. (1983). *The semiology of graphics.* Madison, Wisconsin: University of Wisconsin Press.

Cardinaels, E. (2008). The interplay between cost accounting knowledge and presentation formats in cost-based decision-making. *Accounting, Organizations and Society, 33*(6), 582-602.

Chernoff, H. (1973). The use of faces to represent points in k-dimensional sapce graphically. *Journal of the American Statistical Association, 68*(342), 361-368.

Cleveland, W. S., & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association, 79*(387), 531-554.

Davis, L. R. (1989). Report format and the decision makers task: An experimental investigation. *Accounting, Organizations and Society, 14*(5/6), 495-508.

Desanctis, G., & Jarvenpaa, S. L. (1989). Graphical presentation of accounting data for financial forecasting: An experimental investigation. *Accounting, Organizations and Society, 14*(5/6), 509-525.

Dooley, R. P., & Harkins, L. E. (1970). Functional and attention getting effects of color on graphic communication. *Perceptual and Motor Skills, 31*, 851-854.

Frownfelter-Lohrke, C. (1998). The effects of differing information presentations of general purpose financial statements on users' decisions. *Journal of Information Systems, 12*(2), 99-107.

Harvey, N., & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgmental forecasting. *International Journal of Forecasting, 12*(1), 119-137.

Hill, W. Y., & Milner, M. (2003). Guidelines for graphical displays in financial reporting. *Accounting Education, 12*(2), 531-554.

Hwang, M. I. (1995). The effectiveness of graphic and tabular presentation under time pressure and task complexity. *Information Resource Management Journal, 8*(3), 25-31.

IFRS Foundation. (2010). Staff Draft of Exposure Draft: Financial Statement Presentation. London: IFRS Foundation. Retrieved from http://www.fondazioneoic.eu/wp-content/uploads/downloads/2010/12/2010-07_Staff-draft-FSP_Standard.pdf

Lucas, H. C. (1981). An experimental investigation of the use of computer-based graphics in decision making. *Management Science, 27*(7), 757-768.

Mackay, D. B., & Villarreal, A. (1987). Performance differences in the use of graphic and tabular displays of multivariate data. *Decision Sciences, 18*(4), 535-546.

Masnick, A. M., & Tinkelman, D. (2015, August 8). *Can Users Perceive Minimally Material Weekly Patterns and Anomalies in Embedded Income Statement Sparklines?* Retrieved March 20,

2016, from Social Science Research Network: http://papers.ssrn.com.eur.idm.oclc.org/sol3/papers.cfm?abstract_id=2641302

Parsons, L. M., & Tinkelman, D. (2013). Testing the feasibility of small multiples of sparklines to display semimonthly income statement data. *International Journal of Accounting Information Systems, 14*(1), 58-76.

So, S., & Smith, M. (2002). Colour graphics and task complexity in multivariate decision-making. *Accounting, Auditing and Accountability Journal, 15*(4), 565-592.

So, S., & Smith, M. (2004). Multivariate decision accuracy and the presentation of accounting information. *Accounting Forum, 28*(3), 283-305.

Stock, D., & Watson, C. J. (1984). Human judgment accuracy, multidimensional graphics, and human versus models. *Journal of Accounting Research, 22*(1), 192-206.

Sullivan, J. J. (1988). Financial presentation format and managerial decision making: Tables versus graphics. *Management, 2*(2), 194-216.

Tufte, E. R. (2006). Sparklines: Intense, Simple, Word-Sized Graphics. In E. R. Tufte, *Beautiful Evidence* (pp. 47-63). Cheshire (Connecticut): Graphics Press.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences, 22*(2), 219-240.

9 Appendix

9.1 Experiment

Figure 1: Survey Flow

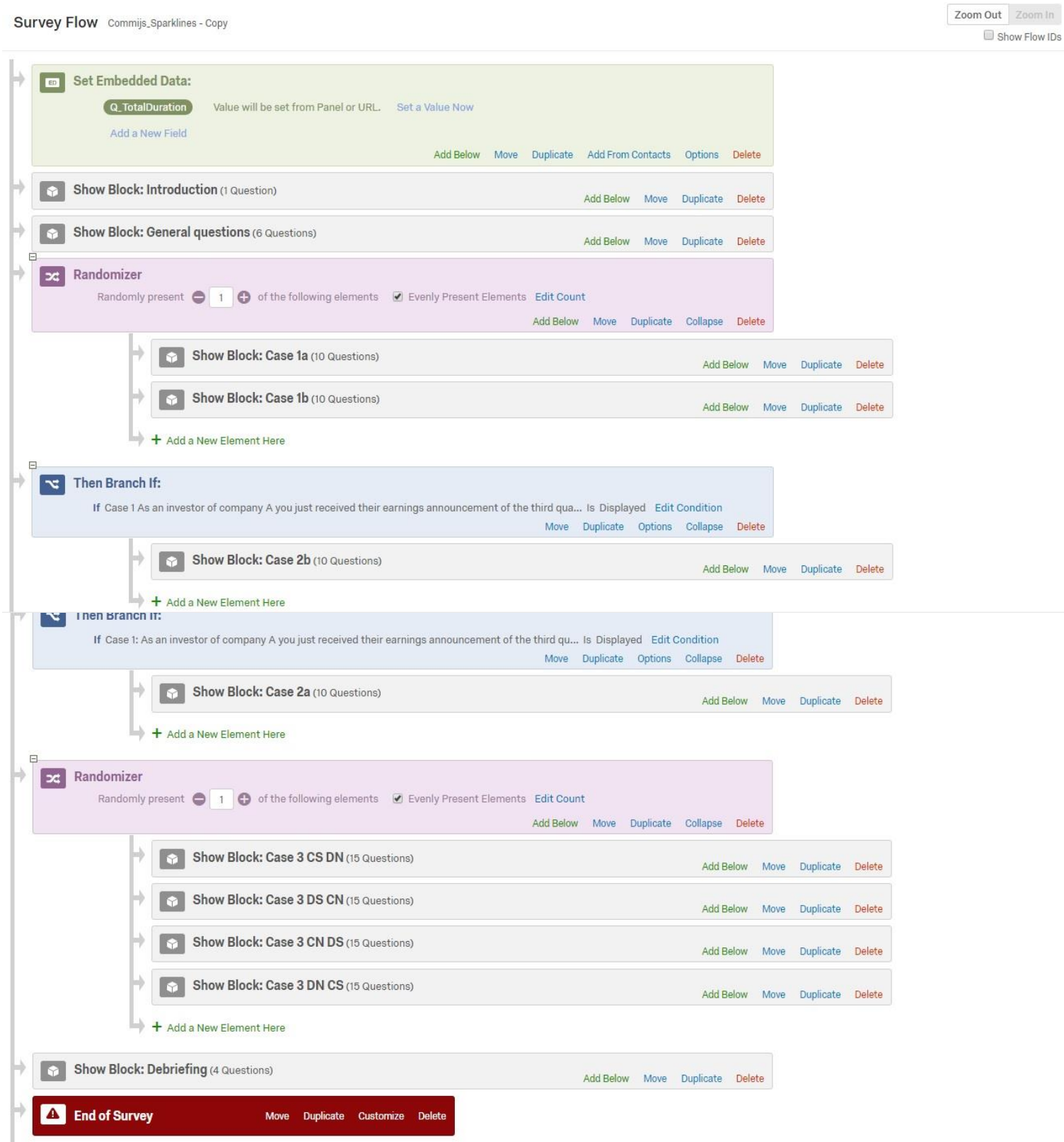


Figure 2: Introduction experiment

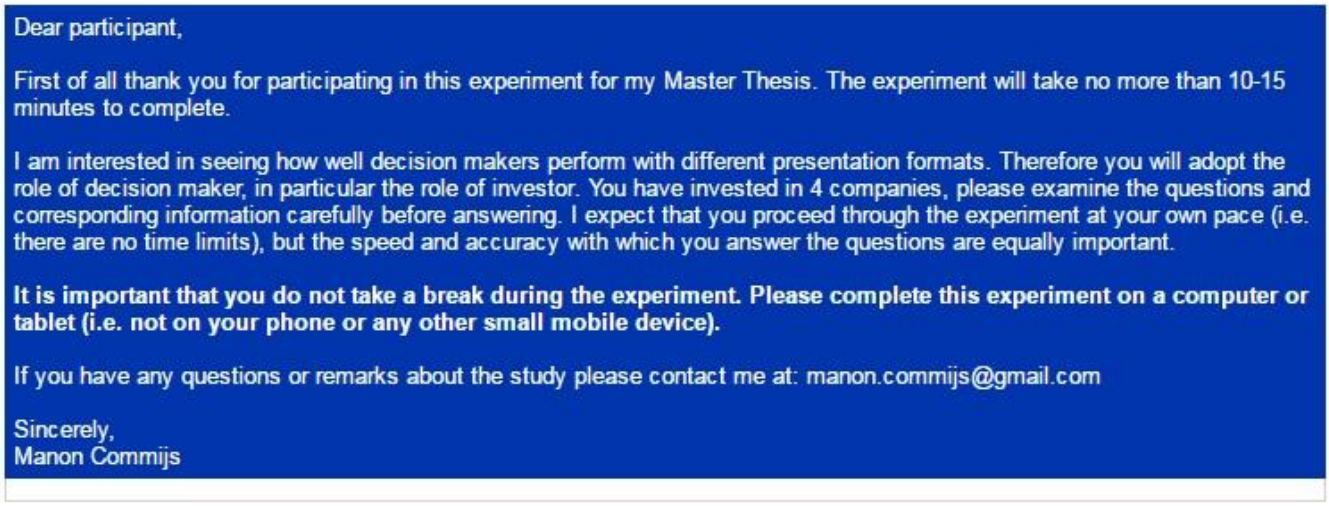


Figure 3: General questions

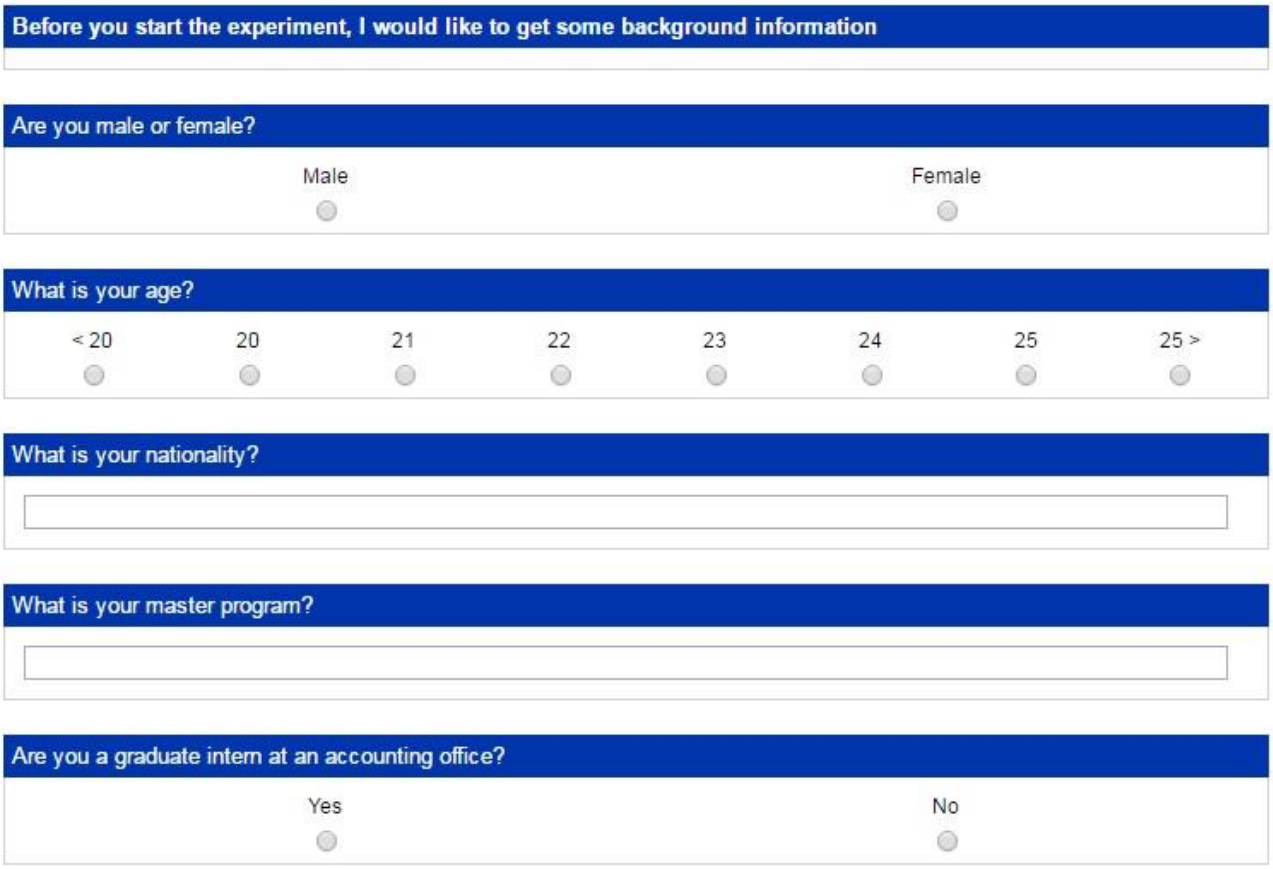


Figure 4: Company A with and without sparklines

Company A: Consolidated income statement (in millions of euros)						
	2014			2015		
	Q1	Q2	Q3	Q1	Q2	Q3
Sales	9.384	9.938	10.388	10.678	9.948	8.672
Cost of sales	5.584	5.988	6.984	6.446	5.958	5.128
Gross margin	3.800	3.950	3.404	4.232	3.990	3.544
Selling expenses	2.332	2.428	2.490	2.682	2.480	2.280
General and administrative expenses	334	352	382	428	408	392
Research and development expenses	792	800	744	872	866	852
Other business income	20	18	42	44	40	33
Other business expenses	12	6	108	16	15	4
Income from operations	350	382	(278)	278	261	49
Financial income	32	30	128	62	56	24
Financial expenses	170	178	288	196	204	224
Income before taxes	212	234	(438)	144	113	(151)
Income taxes (loss)	(56)	(64)	100	(62)	(56)	(16)
Income after taxes	156	170	(338)	82	57	(167)
Results Investments in associates (loss)	42	6	78	46	(2)	4
Net income from continuing operations	198	176	(260)	128	55	(163)
Discontinued operations	82	110	54	72	136	176
Net income	280	286	(206)	200	191	13

Company A: Consolidated income statement (in millions of euros)							
	2014				2015			
	Q1	Q2	Q3		Q1	Q2	Q3	
Sales	9.384	9.938	10.388		10.678	9.948	8.672	
Cost of sales	5.584	5.988	6.984		6.446	5.958	5.128	
Gross margin	3.800	3.950	3.404		4.232	3.990	3.544	
Selling expenses	2.332	2.428	2.490		2.682	2.480	2.280	
General and administrative expenses	334	352	382		428	408	392	
Research and development expenses	792	800	744		872	866	852	
Other business income	20	18	42		44	40	33	
Other business expenses	12	6	108		16	15	4	
Income from operations	350	382	(278)		278	261	49	
Financial income	32	30	128		62	56	24	
Financial expenses	170	178	288		196	204	224	
Income before taxes	212	234	(438)		144	113	(151)	
Income taxes (loss)	(56)	(64)	100		(62)	(56)	(16)	
Income after taxes	156	170	(338)		82	57	(167)	
Results Investments in associates (loss)	42	6	78		46	(2)	4	
Net income from continuing operations	198	176	(260)		128	55	(163)	
Discontinued operations	82	110	54		72	136	176	
Net income	280	286	(206)		200	191	13	

Figure 5: Company A additional information with and without sparklines

Company A: Consolidated income statement (in millions of euros)				
	2014			
	Q1	Q2	Q3	Q4
Sales	9.384	9.938	10.388	13.072
Cost of sales	5.584	5.988	6.984	8.014
Gross margin	3.800	3.950	3.404	5.058
Selling expenses	2.332	2.428	2.490	2.998
General and administrative expenses	334	352	382	426
Research and development expenses	792	800	744	934
Other business income	20	18	42	46
Other business expenses	12	6	108	422
Income from operations	350	382	(278)	324
Financial income	32	30	128	38
Financial expenses	170	178	288	194
Income before taxes	212	234	(438)	168
Income taxes (loss)	(56)	(64)	100	(32)
Income after taxes	156	170	(338)	136
Results Investments in associates (loss)	42	6	78	(2)
Net income from continuing operations	198	176	(260)	134
Discontinued operations	82	110	54	134
Net income	280	286	(206)	268

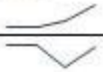
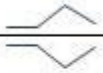
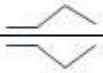
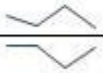
Company A: Consolidated income statement (in millions of euros)					
	2014				
	Q1	Q2	Q3	Q4	
Sales	9.384	9.938	10.388	13.072	
Cost of sales	5.584	5.988	6.984	8.014	
Gross margin	3.800	3.950	3.404	5.058	
Selling expenses	2.332	2.428	2.490	2.998	
General and administrative expenses	334	352	382	426	
Research and development expenses	792	800	744	934	
Other business income	20	18	42	46	
Other business expenses	12	6	108	422	
Income from operations	350	382	(278)	324	
Financial income	32	30	128	38	
Financial expenses	170	178	288	194	
Income before taxes	212	234	(438)	168	
Income taxes (loss)	(56)	(64)	100	(32)	
Income after taxes	156	170	(338)	136	
Results Investments in associates (loss)	42	6	78	(2)	
Net income from continuing operations	198	176	(260)	134	
Discontinued operations	82	110	54	134	
Net income	280	286	(206)	268	

Figure 6: Company B with and without sparklines

Company B: Consolidated income statement (in millions)						
	2014			2015		
	Q1	Q2	Q3	Q1	Q2	Q3
Revenue	8.479	16.227	8.307	9.606	18.734	9.762
Cost of sales	3.590	6.848	3.423	3.982	7.702	3.908
Operating expenses	2.580	5.071	2.594	2.890	5.806	3.048
Depreciation and amortization expenses	339	688	360	412	846	473
General and administrative expenses	485	966	539	597	1.209	577
Total operating expenses	6.994	13.573	6.916	7.881	15.563	8.006
Income from equity investees	102	222	146	106	215	121
Operating income	1.587	2.876	1.537	1.831	3.386	1.877
Interest income	40	75	39	19	22	51
Interest expense	29	63	33	33	66	38
Earnings before income taxes	1.598	2.888	1.543	1.817	3.342	1.890
Income taxes (loss)	(556)	(993)	(518)	(630)	(1.162)	(637)
Net earnings incl. noncontrolling interests	1.042	1.895	1.025	1.187	2.180	1.253
Net earnings attributable to noncontrolling interests	-	-	0,2	(3)	(4)	0,4
Net earnings	1.042	1.895	1.025	1.184	2.176	1.253

Company B: Consolidated income statement (in millions)							
	2014				2015			
	Q1	Q2	Q3		Q1	Q2	Q3	
Revenue	8.479	16.227	8.307		9.606	18.734	9.762	
Cost of sales	3.590	6.848	3.423		3.982	7.702	3.908	
Operating expenses	2.580	5.071	2.594		2.890	5.806	3.048	
Depreciation and amortization expenses	339	688	360		412	846	473	
General and administrative expenses	485	966	539		597	1.209	577	
Total operating expenses	6.994	13.573	6.916		7.881	15.563	8.006	
Income from equity investees	102	222	146		106	215	121	
Operating income	1.587	2.876	1.537		1.831	3.386	1.877	
Interest income	40	75	39		19	22	51	
Interest expense	29	63	33		33	66	38	
Earnings before income taxes	1.598	2.888	1.543		1.817	3.342	1.890	
Income taxes (loss)	(556)	(993)	(518)		(630)	(1.162)	(637)	
Net earnings incl. noncontrolling interests	1.042	1.895	1.025		1.187	2.180	1.253	
Net earnings attributable to noncontrolling interests	-	-	0,2		(3)	(4)	0,4	
Net earnings	1.042	1.895	1.025		1.184	2.176	1.253	

Figure 7: Company B additional information with and without sparklines

Company B: Consolidated income statement (in millions)				
	2014			
	Q1	Q2	Q3	Q4
Revenue	8.479	16.227	8.307	8.362
Cost of sales	3.590	6.848	3.423	3.446
Operating expenses	2.580	5.071	2.594	2.526
Depreciation and amortization expenses	339	688	360	371
General and administrative expenses	485	966	539	477
Total operating expenses	6.994	13.573	6.916	6.820
Income from equity investees	102	222	146	169
Operating income	1.587	2.876	1.537	1.711
Interest income	40	75	39	171
Interest expense	29	63	33	33
Earnings before income taxes	1.598	2.888	1.543	1.849
Income taxes (loss)	(556)	(993)	(518)	(673)
Net earnings incl. noncontrolling interests	1.042	1.895	1.025	1.176
Net earnings attributable to noncontrolling interests	-	-	0,2	0,6
Net earnings	1.042	1.895	1.025	1.177

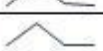
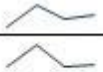
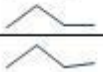
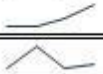
Company B: Consolidated income statement (in millions)					
	2014				
	Q1	Q2	Q3	Q4	
Revenue	8.479	16.227	8.307	8.362	
Cost of sales	3.590	6.848	3.423	3.446	
Operating expenses	2.580	5.071	2.594	2.526	
Depreciation and amortization expenses	339	688	360	371	
General and administrative expenses	485	966	539	477	
Total operating expenses	6.994	13.573	6.916	6.820	
Income from equity investees	102	222	146	169	
Operating income	1.587	2.876	1.537	1.711	
Interest income	40	75	39	171	
Interest expense	29	63	33	33	
Earnings before income taxes	1.598	2.888	1.543	1.849	
Income taxes (loss)	(556)	(993)	(518)	(673)	
Net earnings incl. noncontrolling interests	1.042	1.895	1.025	1.176	
Net earnings attributable to noncontrolling interests	-	-	0,2	0,6	
Net earnings	1.042	1.895	1.025	1.177	

Figure 8: Company C and D with and without sparklines

Company C: Consolidated income statement (in millions)	2014			2015		
	Q1	Q2	Q3	Q1	Q2	Q3
Sales	5.554	6.351	6.695	5.656	6.642	7.415
External expenses	3.839	3.985	4.246	4.038	4.382	4.355
Salaries and related costs	1.837	2.015	1.850	1.876	1.965	1.888
Amortization, depreciation and provisions	395	403	435	426	384	450
Other income and expenses	118	128	80	388	310	218
Income from operating activities	(399)	76	244	(296)	221	940
Foreign exchange gains (loss), net	(117)	(2)	(59)	(155)	(91)	(75)
Income before taxes	(516)	74	185	(451)	130	865
Income taxes	69	(33)	(15)	36	46	(96)
Net income	(447)	41	170	(415)	176	769

Company D: Consolidated income statement (in millions of euros)	2014			2015		
	Q1	Q2	Q3	Q1	Q2	Q3
Revenue	3.970	4.268	4.516	4.431	4.816	5.097
Cost of services	3.252	3.481	3.681	3.617	3.913	4.043
Gross profit	718	787	835	814	903	1.054
Selling expenses	409	422	427	459	473	486
General and administrative expenses	190	196	204	202	204	206
Operating profit (loss)	119	169	204	153	226	362
Net finance profit (loss)	(3)	(4)	(11)	(22)	(4)	(6)
Income before taxes	116	165	193	131	222	356
Taxes on income	(25)	(39)	(43)	(22)	(45)	(57)
Net income	91	126	150	109	177	299

Company C: Consolidated income statement (in millions)	2014			2015		
	Q1	Q2	Q3	Q1	Q2	Q3
Sales	5.554	6.351	6.695	5.656	6.642	7.415
External expenses	3.839	3.985	4.246	4.038	4.382	4.355
Salaries and related costs	1.837	2.015	1.850	1.876	1.965	1.888
Amortization, depreciation and provisi	395	403	435	426	384	450
Other income and expenses	118	128	80	388	310	218
Income from operating activities	(399)	76	244	(296)	221	940
Foreign exchange gains (loss), net	(117)	(2)	(59)	(155)	(91)	(75)
Income before taxes	(516)	74	185	(451)	130	865
Income taxes	69	(33)	(15)	36	46	(96)
Net income	(447)	41	170	(415)	176	769

Company D: Consolidated income statement (in millions of euros)	2014			2015		
	Q1	Q2	Q3	Q1	Q2	Q3
Revenue	3.970	4.268	4.516	4.431	4.816	5.097
Cost of services	3.252	3.481	3.681	3.617	3.913	4.043
Gross profit	718	787	835	814	903	1.054
Selling expenses	409	422	427	459	473	486
General and administrative expenses	190	196	204	202	204	206
Operating profit (loss)	119	169	204	153	226	362
Net finance profit (loss)	(3)	(4)	(11)	(22)	(4)	(6)
Income before taxes	116	165	193	131	222	356
Taxes on income	(25)	(39)	(43)	(22)	(45)	(57)
Net income	91	126	150	109	177	299

Company D: Consolidated income statement (in millions of euros)	2014			2015		
	Q1	Q2	Q3	Q1	Q2	Q3
Revenue	3.970	4.268	4.516	4.431	4.816	5.097
Cost of services	3.252	3.481	3.681	3.617	3.913	4.043
Gross profit	718	787	835	814	903	1.054
Selling expenses	409	422	427	459	473	486
General and administrative expenses	190	196	204	202	204	206
Operating profit (loss)	119	169	204	153	226	362
Net finance profit (loss)	(3)	(4)	(11)	(22)	(4)	(6)
Income before taxes	116	165	193	131	222	356
Taxes on income	(25)	(39)	(43)	(22)	(45)	(57)
Net income	91	126	150	109	177	299

Company C: Consolidated income statement (in millions)	2014			2015		
	Q1	Q2	Q3	Q1	Q2	Q3
Sales	5.554	6.351	6.695	5.656	6.642	7.415
External expenses	3.839	3.985	4.246	4.038	4.382	4.355
Salaries and related costs	1.837	2.015	1.850	1.876	1.965	1.888
Amortization, depreciation and provisi	395	403	435	426	384	450
Other income and expenses	118	128	80	388	310	218
Income from operating activities	(399)	76	244	(296)	221	940
Foreign exchange gains (loss), net	(117)	(2)	(59)	(155)	(91)	(75)
Income before taxes	(516)	74	185	(451)	130	865
Income taxes	69	(33)	(15)	36	46	(96)
Net income	(447)	41	170	(415)	176	769

Company C: Consolidated income statement (in millions)	2014			2015		
	Q1	Q2	Q3	Q1	Q2	Q3
Sales	5.554	6.351	6.695	5.656	6.642	7.415
External expenses	3.839	3.985	4.246	4.038	4.382	4.355
Salaries and related costs	1.837	2.015	1.850	1.876	1.965	1.888
Amortization, depreciation and provisions	395	403	435	426	384	450
Other income and expenses	118	128	80	388	310	218
Income from operating activities	(399)	76	244	(296)	221	940
Foreign exchange gains (loss), net	(117)	(2)	(59)	(155)	(91)	(75)
Income before taxes	(516)	74	185	(451)	130	865
Income taxes	69	(33)	(15)	36	46	(96)
Net income	(447)	41	170	(415)	176	769

Figure 9: Questions for part 1

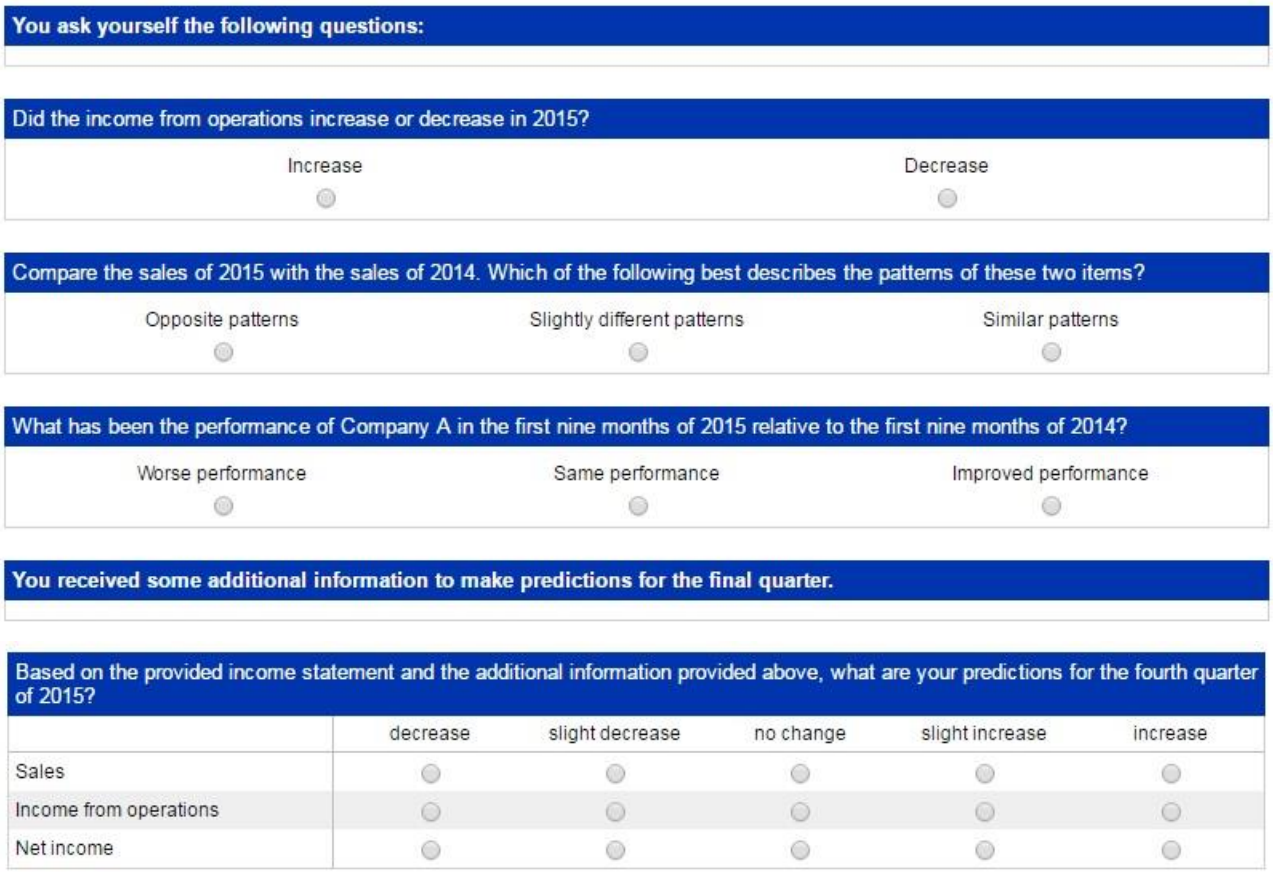


Figure 10: Questions for part 2, same questions for company D as company C

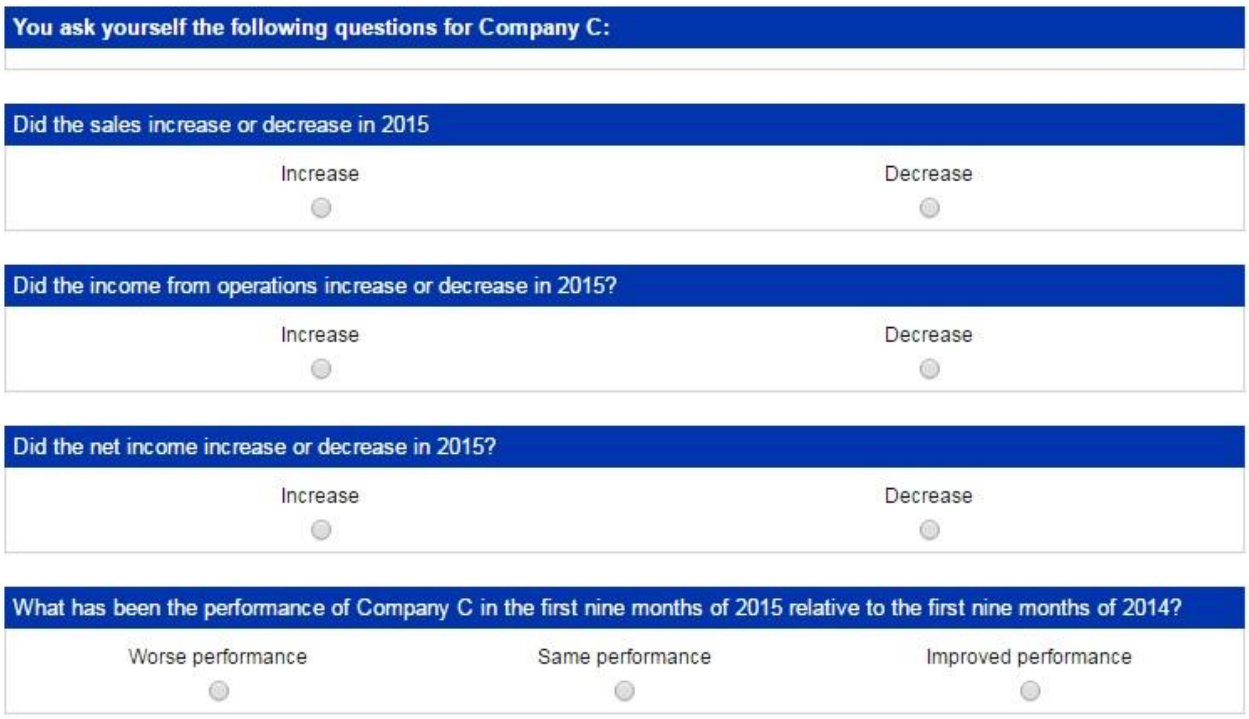
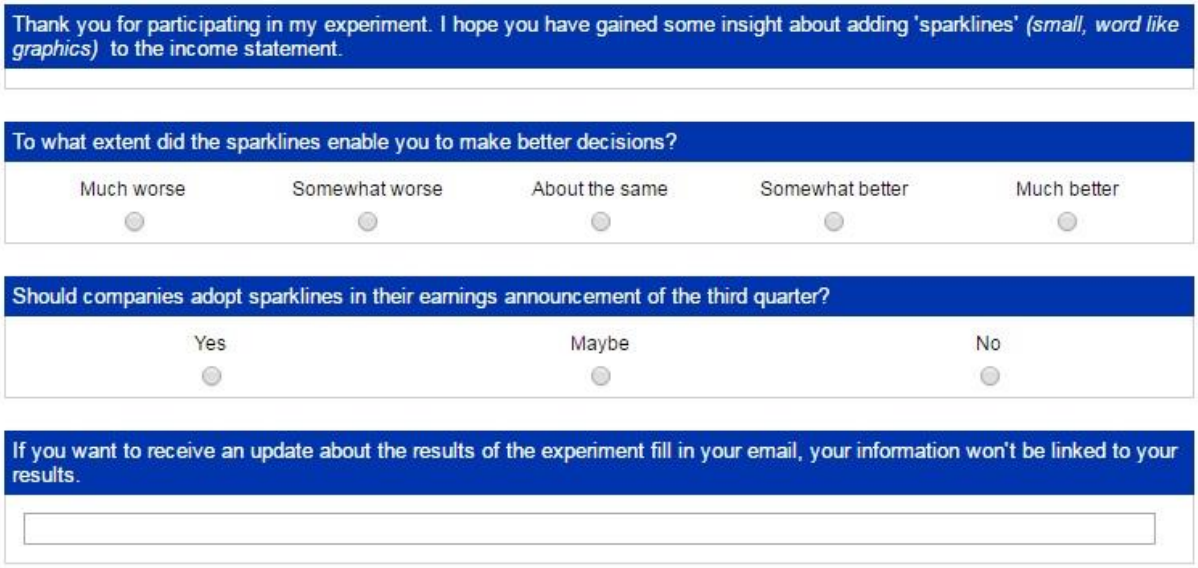
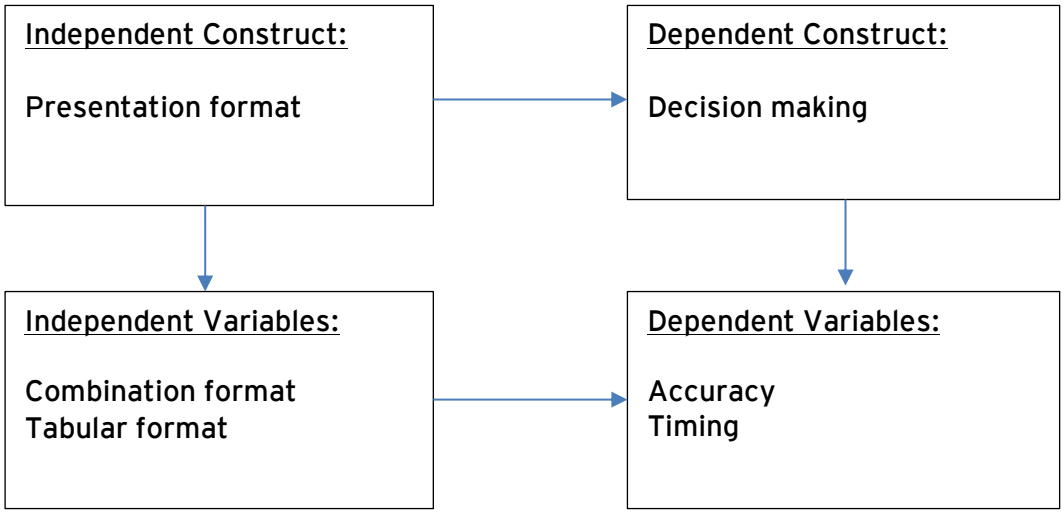


Figure 11: Debriefing



9.2 Libby Boxes

Figure 12: Libby Boxes



9.3 Subjects

Figure 13: Age distribution

	21	22	23	24	25	25 >	Total
Male	2	5	3	3	3	7	**23**
Female	1	5	6	1	0	4	**17**
Total	**3**	**10**	**9**	**4**	**3**	**11**	**40**

Figure 14: Nationality distribution

	Nationality		
	Other	**Dutch**	**Total**
Male	2	21	**23**
Female	5	12	**17**
Total	**7**	**33**	**40**

Figure 15: Internship distribution

	Graduate intern		
	Yes	**No**	**Total**
Male	14	9	**23**
Female	6	11	**17**
Total	**20**	**20**	**40**

9.4 Additional analysis responses – Before elimination

Figure 16: Timing Case 1A & 2B

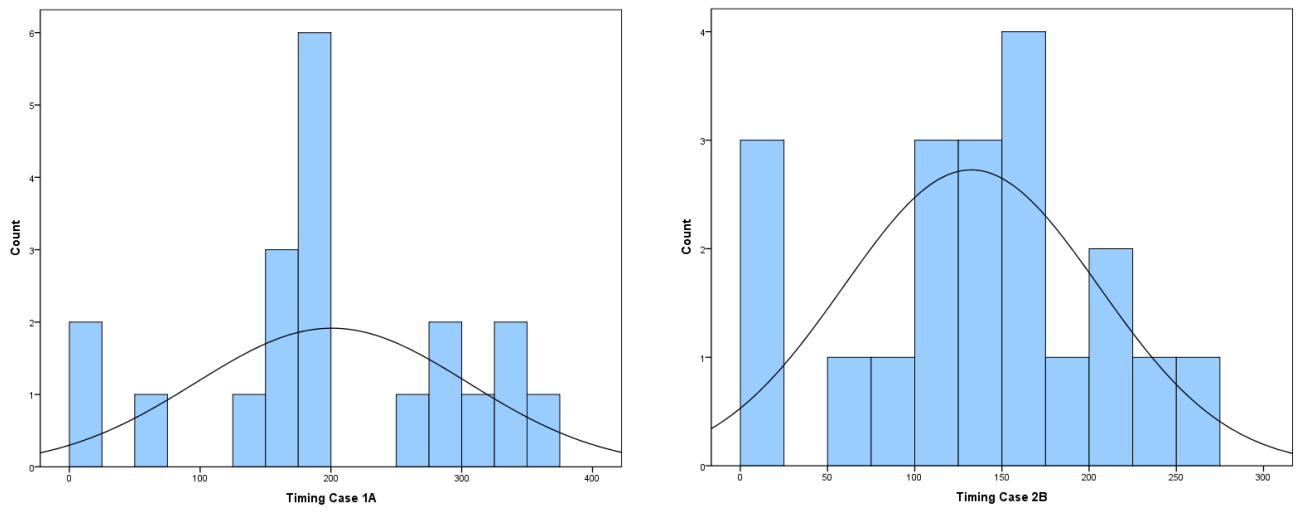
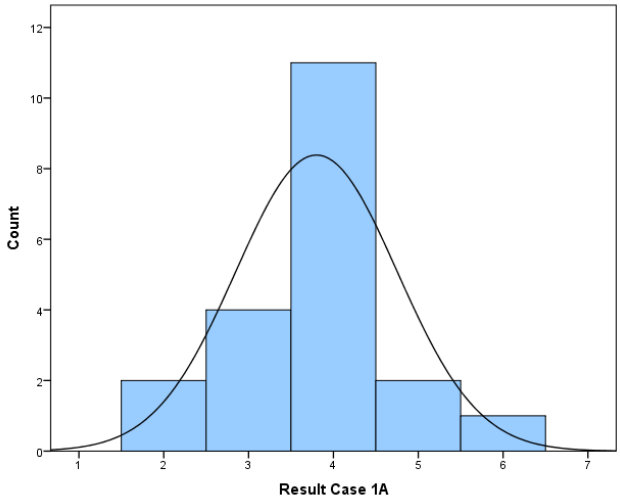


Figure 17: Result Case 1A & 2B



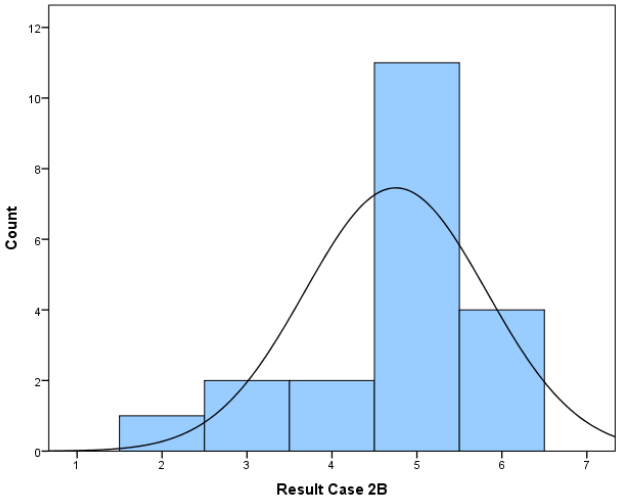


Figure 18: Total Result Case 1A & 2B

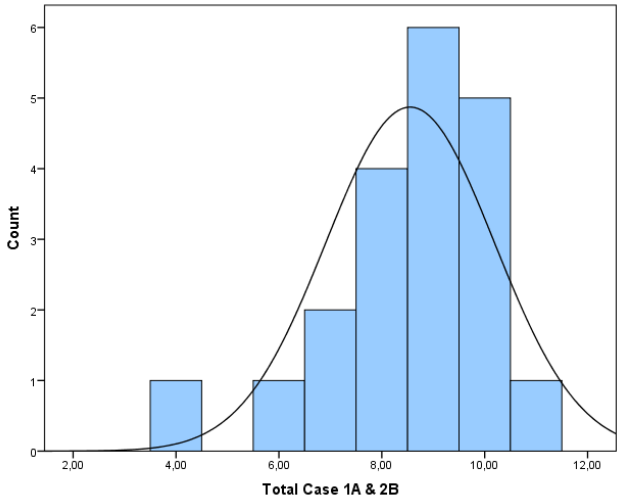
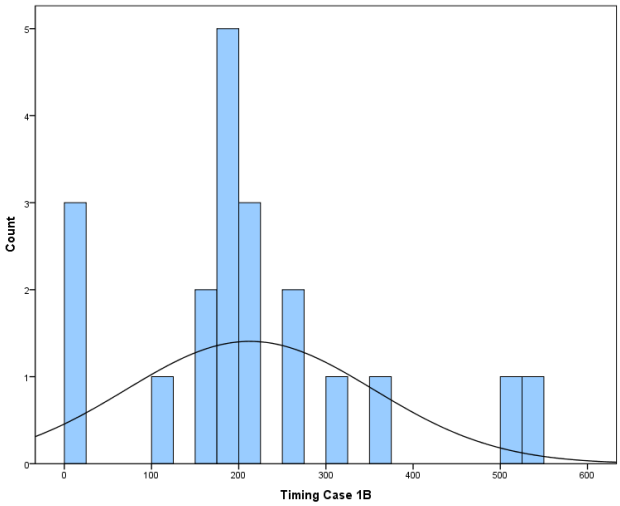


Figure 19: Timing Case 1B & 2A



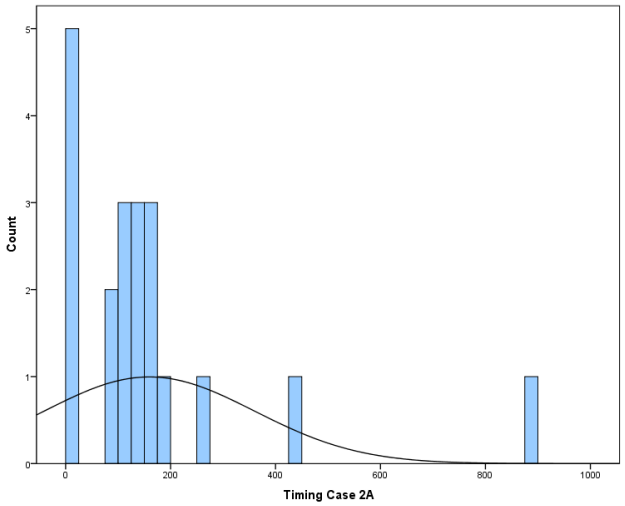
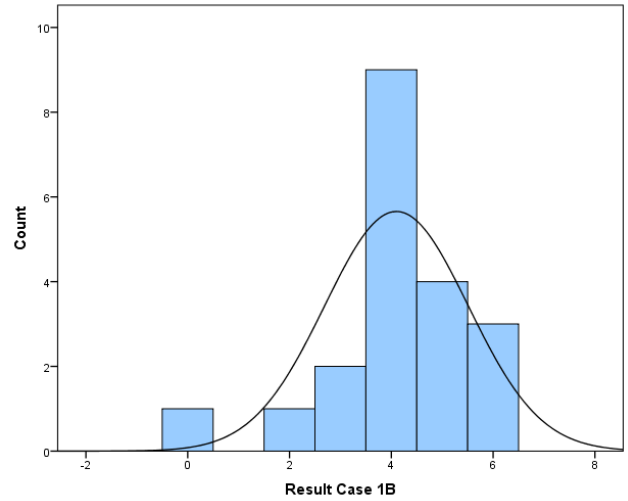


Figure 20: Result Case 1B & 2A



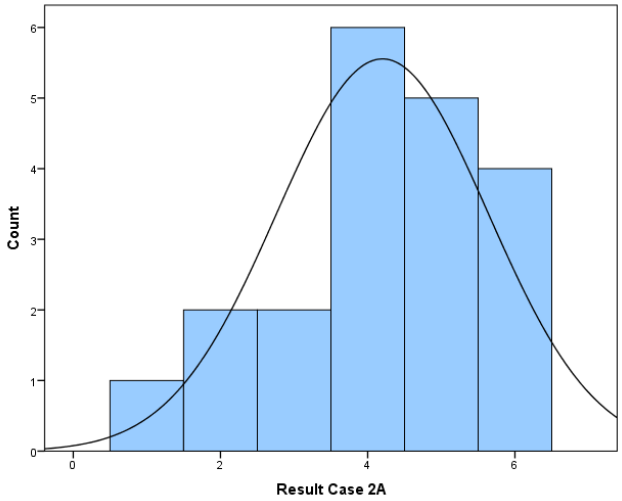


Figure 21: Total Result Case 1B & 2A

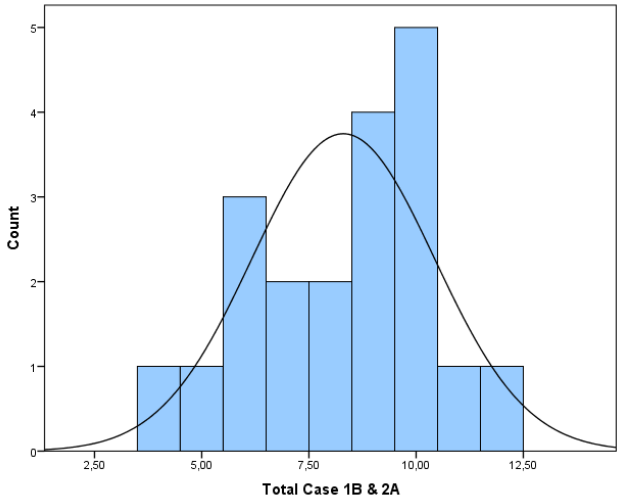
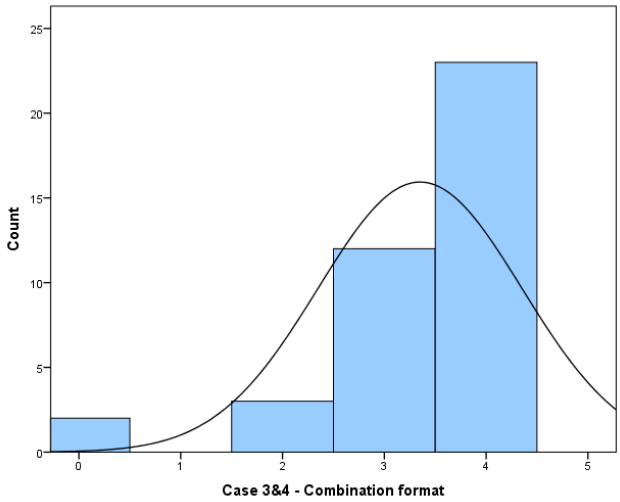


Figure 22: Result Case 3&4



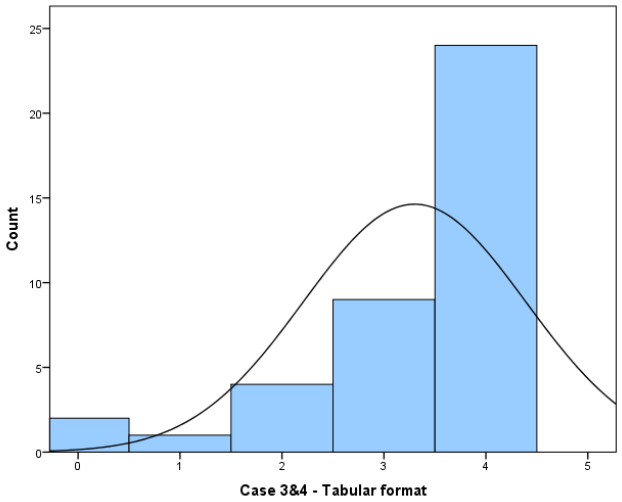


Figure 23: Elimination responses

	Result Case 1A	Timing Case 1A	Result Case 2B	Timing Case 2B	Total	Result Case 3&4 - Combination format	Result Case 3&4 - Tabular format	Total
R_1GCytghiGnZ5453	4	185,886	5	160,945	9	4	4	8
R_1Is0QnUN63xSZ9J	4	374,995	6	192,926	10	4	4	8
R_1jddZsgr70h1s2I	4	345,403	6	215,291	10	4	4	8
R_1NmP4Zg3n3wXogE	4	135,926	3	74,006	7	4	0	4
R_1rGDSMT9iEJjZEg	4	307,743	5	237,982	9	3	4	7
R_24IACB9WotpnC5w	4	275,059	6	210,864	10	4	4	8
R_2aghJ7Y3sWvZsKN	2	0,211	4	121,626	6	3	3	6
R_2VKZ0q4K6o6aSm7	5	165,849	5	129,933	10	3	3	6
R_32QHswx0tbonSnc	4	197,032	4	145,326	8	3	1	4
R_3e2dE6Iqe2wSlbZ	3	347,813	5	0,184	8	4	4	8
R_3EWRQWQBb8K5oJX	3	178,474	6	103,596	9	4	4	8
R_3nJzkpcSNYImG8O	4	293,156	5	255,153	9	3	4	7
R_3sstZdFrpvulUKY	4	19,475	5	150,986	9	3	4	7
R_5pThRKuTboN7WOR	3	257,512	5	116,789	8	4	4	8
R_6GB7DdOkVEDSrER	5	178,828	5	157,133	10	3	3	6
R_qDXxkJjR25DoZlT	4	191,509	5	129,736	9	4	4	8
R_r857aMDNb7kh0lz	3	152,102	5	0,082	8	3	4	7
R_RFeTO5VHngwd6md	2	56,483	2	14,556	4	2	2	4
R_vkSZHyGm8ltFrFf	6	184,957	5	152,493	11	4	4	8
R_XNFSdsBcOzLxZm1	4	165,726	3	78,051	7	3	3	6
Grand Total	**76**	**4014,139**	**95**	**2647,658**	**171**	**69**	**67**	**136**

Timing

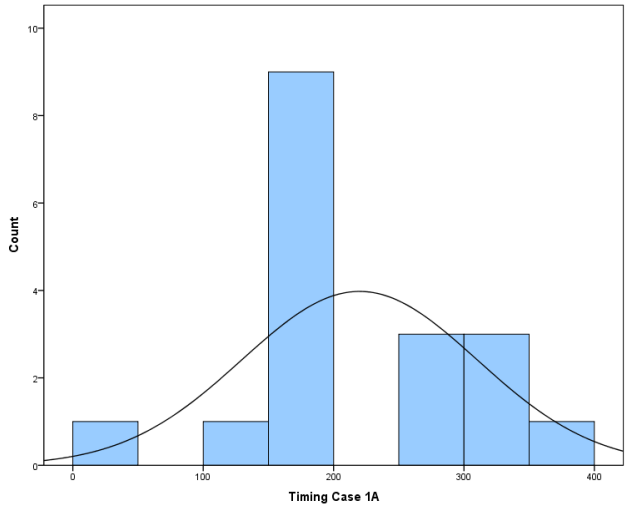
Result

Total Result

Row Labels	Result Case 1A	Timing Case 1A	Result Case 2B	Timing Case 2B	Total	Result Case 3&4 - Combination format	Result Case 3&4 - Tabular format	Total
R_1DAJYsNDG2lHbwv	4	112,585	6	201,948	10	3	3	6
R_1dnz5Y9uGmt3INo	5	137,227	4	210,419	9	4	3	7
R_1f958QMg9C0YtJb	6	267,248	4	506,261	10	4	4	8
R_1g1tPYrYbTgJC6K	5	883,826	4	0,222	9	4	0	4
R_1HnnPCcm2CcpaSW	5	135,647	6	272,867	11	2	2	4
R_1IFcSdYRjojlJbF	4	101,396	4	162,001	8	3	3	6
R_1obdfpj8u4ZtXAe	1	10,713	5	169,066	6	0	2	2
R_1ojVJ4hbmvytD7S	5	158,883	4	260,281	9	4	4	8
R_1zyjJAszx4Ki7Ul	5	0,136	5	314,276	10	4	4	8
R_2f80baB1mmRQdnh	6	76,274	6	124,733	12	4	4	8
R_2frJjfG2QMDZz3C	4	163,728	5	183,716	9	4	4	8
R_3012RfI8x4Gbi8w	2	153,341	3	203,169	5	0	4	4
R_3GrIkqCyU2XzYSt	4	0,268	0	362,828	4	4	4	8
R_3HzxSHApd7srrJc	2	17,009	4	0,216	6	2	2	4
R_3Mg50qkTUD8wiU5	3	194,455	4	199,102	7	4	3	7
R_3PZCyLmIprhe8e9	4	0,110	3	190,655	7	4	4	8
R_6WlBbrm3sXuH8TT	6	124,515	4	178,397	10	4	4	8
R_Q55fBiW6DzCkekx	3	77,796	5	177,157	8	4	3	7
R_u8tG63Zpin6HBDz	6	127,303	4	0,027	10	3	4	7
R_UgukcOdiio99O7f	4	447,922	2	537,377	6	4	4	8
Grand Total	**84**	**3190,382**	**82**	**4254,718**	**166**	**65**	**65**	**130**

9.5 Additional analysis responses – After elimination

Figure 24: Timing Case 1A & 2B



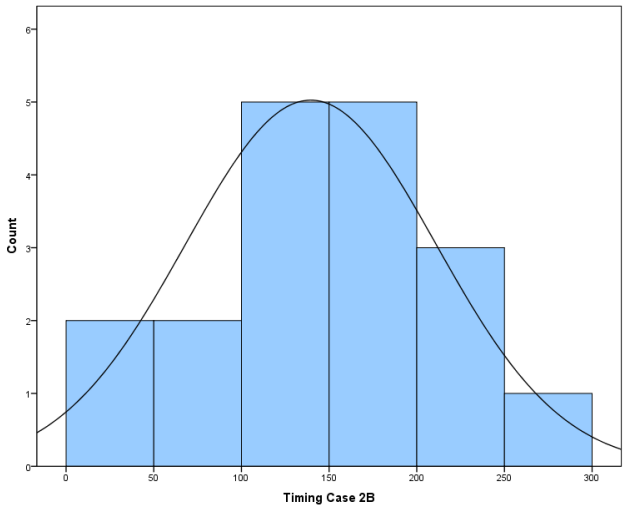
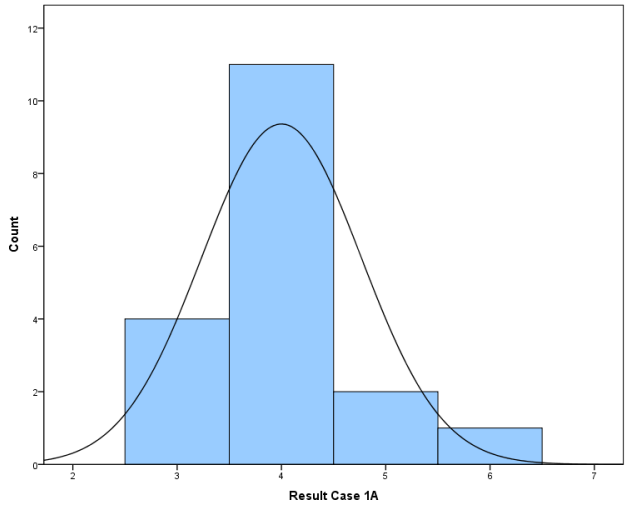


Figure 25: Result Case 1A & 2B



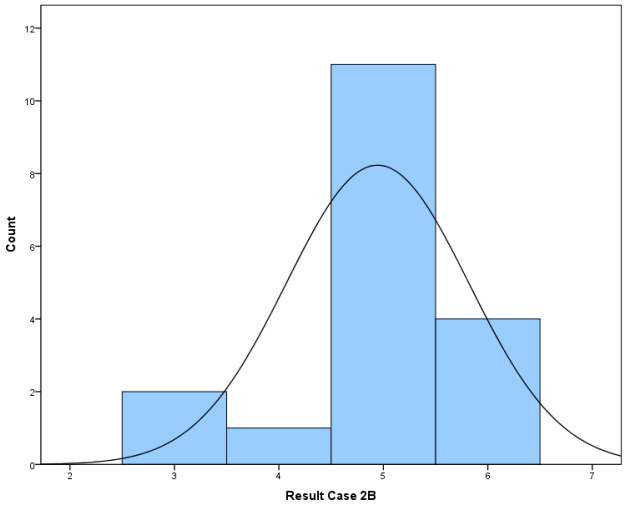


Figure 26: Total Result Case 1A & 2B

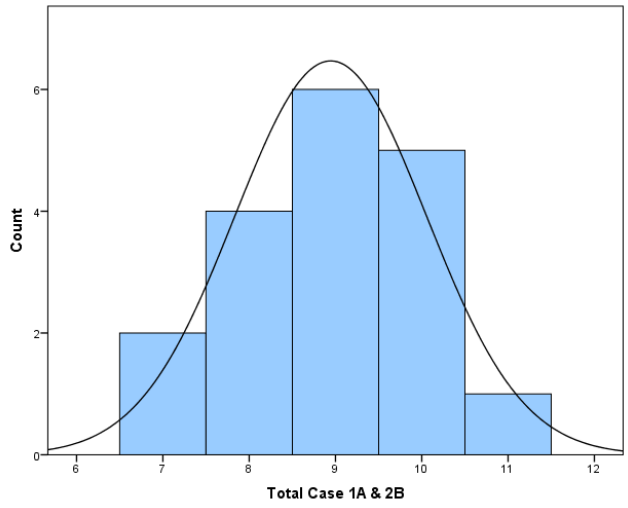
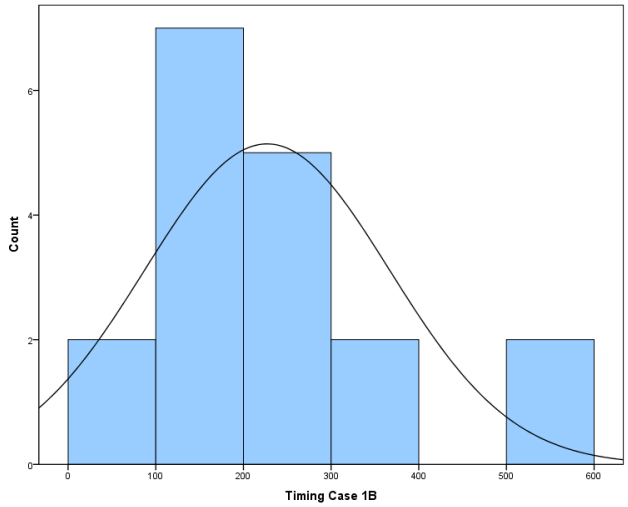


Figure 27: Timing Case 1B & 2A



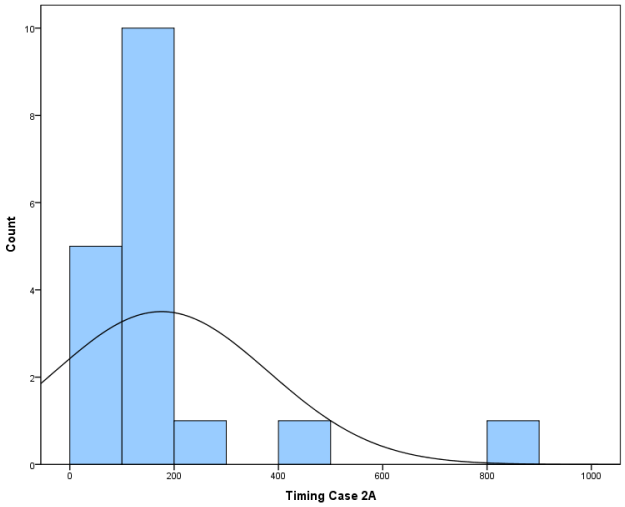
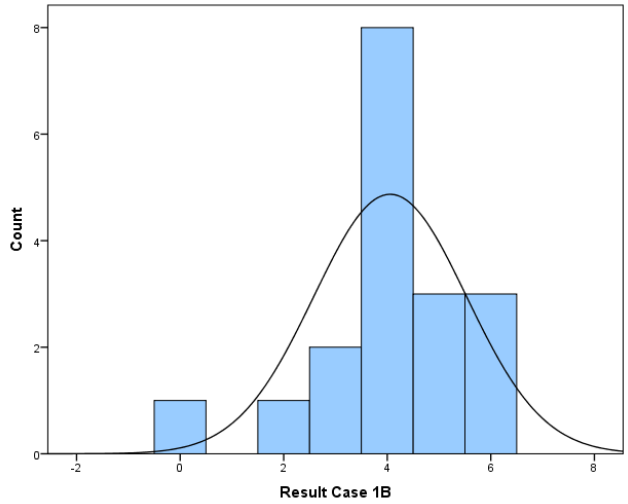


Figure 28: Result Case 1B & 2A



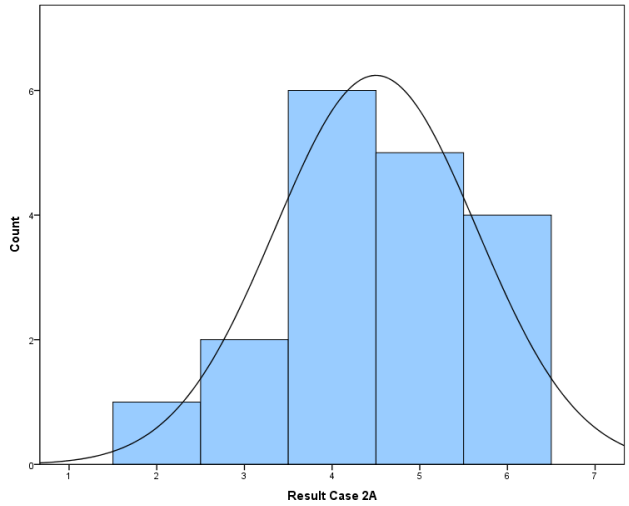


Figure 29: Total Result Case 1B & 2A

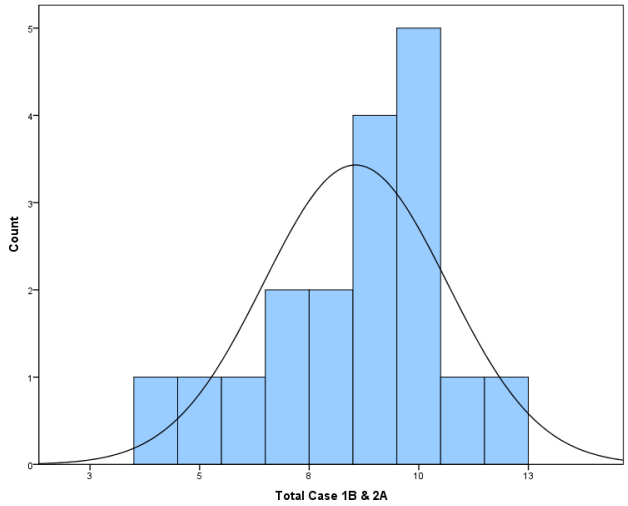


Figure 30: Result Case 3&4