CrossMark

Factors influencing strategy implementation decisions: an evaluation of a balanced scorecard cockpit, intelligence, and knowledge

Jürgen Strohhecker[1]

Abstract Successful strategy implementation is no less important than designing the strategy, but it has long been bemoaned that managerial attention on the implementation stage is low and the failure rate is high. By interpreting strategy implementation as a dynamic decision-making task and building on complex problem solving and performance measurement research, this study investigates whether factors such as a balanced scorecard (BSC) cockpit, intelligence, and knowledge explain differences in strategy implementation performance. Using a computer-based, feedback-rich system dynamics business game, the participants were placed in a top manager position. Their task was to implement a given strategy as best they could, which meant translating strategy into operational decisions over a period of 10 years. One experimental group was equipped with a BSC management cockpit that was designed according to suggestions of the inventors of the concept—Kaplan and Norton—and carefully tailored to the strategy, while the other group had to rely on a traditional report cockpit as an information source. Regression analysis did not yield significant performance differences between the groups. Instead, participants' knowledge and general cognitive intelligence were found to have significant impacts on decision-making performance, supporting findings from previous dynamic decision-making research. This finding is related to the typical BSC cockpit design as originally suggested by Kaplan and Norton, which seems to provide insufficient strategy focus. Finally, avenues for future research are discussed.

✉ Jürgen Strohhecker
j.strohhecker@fs.de

[1] Management Department, Frankfurt School of Finance & Management, Sonnemannstraße 9-11, 60314 Frankfurt, Germany

🙋 Springer

1 Introduction

When following a formal, prescriptive model of strategic management (Ansoff 1984; Dyson 2000; Mintzberg 1990), successful strategy implementation is no less important than designing the strategy. Failed implementation of successful strategies results in lost opportunities of higher organizational performance. However, it has long been bemoaned that managerial attention on the implementation stage is low and the failure rate is high (Alexander 1985; Kaplan and Norton 2004b, 2008; Mankins and Steele 2005; Sterling 2003), although recent research has shown that failure rates of up to 90 % are likely to be overestimations (Cândido and Santos 2015). Modern, balanced performance management concepts—for instance, the balanced scorecard (BSC)— have been increasingly positioned as a remedy for the strategic implementation trap. Over the years, BSC inventors Kaplan and Norton have put increasing emphasis on the scorecard's role as "the cornerstone of a new strategic management process" (1996b, p. 75). They maintain that use of a BSC will increase an organization's ability to execute its strategy and therefore ultimately improve its performance. However, the research topic—to determine the net economic benefit from the use of non-financial measures and the BSC—proposed by Ittner and Larcker (1998) as early as seems still to be lacking a clear answer. While various studies have addressed this issue, mixed outcomes are reported (Biggart et al. 2010; Bryant et al. 2004; Davis and Albright 2004; De Geuser et al. 2009; Hoque and James 2000; Iselin et al. 2008; Ittner et al. 2003; Maiga and Jacobs 2003; Malina and Selto 2001; Sim and Koh 2001; Tapinos et al. 2011).

Nevertheless, based on a comprehensive review of the strategy implementation literature, Atkinson (2006) comes to the conclusion that a BSC can indeed play an important role in strategy implementation when used as a strategic control system (Bungay and Goold 1991) that facilitates translating strategy into decision making and action. Therefore, this study is based on the idea that strategy implementation is a decision-making task. It requires a sequence of multiple and interdependent decisions in a complex dynamic environment—which exactly defines a dynamic decision-making or complex problem-solving task[1] (Brehmer 1992; Edwards 1962). For instance, implementing a growth strategy that relies on superior service quality requires deciding on the number of hires per month. In addition to psychological factors, such as intelligence and knowledge, that are known to affect dynamic decision-making performance, this study focuses specifically on the impact that a scorecard cockpit has on individual managers' dynamic decision making while attempting to "translate strategy into action" (Kaplan and Norton 2004b, p. 52). Such a cockpit assembles and visualizes the key indicators that are included in the BSC metrics—as, for instance, illustrated in (Kaplan and Norton 2001, p. 221). It comprises the adequate "battery of instrumentation" (Kaplan and Norton 1996a, p. 2) that managers need to guide their companies towards future success. The scorecard cockpit displays the measures that have been chosen during the BSC development process in accordance with the company's strategy and four balanced perspectives: financials, customers, internal

---

[1] As both terms are difficult to differentiate in a meaningful way, complex problem solving and dynamic decision making are used as synonyms in this study.

business processes, and learning and growth. Very important to the BSC concept is the fact that the measurement system and strategy are strongly connected (Kaplan and Norton 1996a, p. 148). Measures should never be selected for a BSC cockpit simply because they have already been used in the organization or because they seem to be fashionable; instead, the only reason for them to appear on a scorecard should be that they are the best indicators available for measuring the associated strategic objectives (Kaplan and Norton 1996a, p. 62). Hence, being strongly linked to strategy, a BSC cockpit can be expected to contribute to addressing well-known strategy implementation issues, such as communication, middle-management issues, unclear priorities and targets, undefined actions, insufficient coordination, and inadequate performance monitoring (Atkinson 2006).

Software vendors, consultants, and practically focused books have been actively promoting scorecard cockpits, dashboards, or war rooms for several years (e.g., Alexander 2007; Daum 2006; Eckerson 2011; Person 2008). There, it is implicitly assumed or even advertised that these instruments are important decision support tools worth investing quite considerable amounts of money in. However, beyond anecdotal support, there is a lack of substantial evidence to support this claim. Empirical research that challenges vendors' and consultants' sales arguments and investigates the impact of such cockpits on decision-making performance in general is absent. More specifically, to the best of my knowledge, no study addresses a BSC cockpit's effect on strategy implementation success while also taking important psychological factors into account—intelligence and knowledge.

There is ample evidence that human performance in complex problem solving is generally poor (e.g., Dörner 1980; Dörner et al. 1994; Moxnes 1998; Reichert and Dörner 1988; Sterman 1989; Strohhecker and Größler 2013; Wittmann and Hattrup 2004). Merely implementing a strategy through a series of operational decisions might be a less complex problem than achieving both—developing a potentially successful strategy and successfully implementing it. Nevertheless, it is still a severe challenge for human decision makers (Lane 1999), and failure rates are high (Strohhecker and Größler 2012). Psychological research indicates that general cognitive ability (G) and knowledge (K) relate to dynamic decision-making performance (Ackerman 1996). Therefore, it could be expected that these two factors also affect decision makers' performance in the more operational setting of implementing strategies, which is not investigated in traditional dynamic decision-making research.

This study aims to contribute to both dynamic decision-making research and strategy implementation and empirical BSC research by answering the following research question: Do intelligence, knowledge, and a BSC cockpit, which is designed according to principles and examples described in the BSC literature and closely linked with a well-defined strategy, really improve strategy-implementing decision-making performance by bounded-rational individual decision makers?

As a research method, the laboratory experiment is used following psychological and dynamic decision-making research (Brehmer and Dörner 1993; Capelo and Dias 2009; Lipe and Salterio 2000; Tayler 2010). This has the important advantage that the myriad confounding variables that can substantively impact any results from a field study can be controlled (Sprinkle and Williamson 2007). Most importantly, by using a completely deterministic business simulation without any random variables in the lab-

oratory, potentially broken cause-and-effect links between decision making, decision implementation, and performance results are eliminated. Therefore, financial performance indicators, such as economic value added (EVA), are the undistorted result of the actions taken. In addition, these actions are direct and unbiased implementations of the decisions made, as implementation requires from the individual participants no more than entering some numbers and pressing a button.

The remainder of the paper is organized as follows. The subsequent section reviews related literature and develops the research hypotheses. Section three outlines the research method and design of the laboratory experiment used to test the hypotheses. The results of the statistical analysis of experimental data are presented in section four, followed by a discussion in section five. The paper concludes by highlighting its contribution, discussing some limitations, and providing directions for further research.

2 Theory and hypotheses

Kaplan and Norton's BSC concept has evolved over time from a performance measurement system to a strategic management instrument (Kaplan and Norton 1992, 1996a, 2001, 2004b, 2006, 2008). From a present-day perspective, it consists of two major components. The first is a balanced performance measurement system with a comprehensible number of indicators allocated to four perspectives (Kaplan and Norton 1992, 1996a) recommend between four and seven measures per perspective, and between 16 and 25 measures for the whole scorecard. The four perspectives are as follows: financial perspective, customer perspective, internal process perspective, and learning and growth perspective. The second component is a strategy map that describes the organization's strategy by highlighting and visualizing the cause-and-effect relationships between the strategy's major components (Kaplan and Norton 2004b). Both components have to be closely linked in the sense that measures should clearly operationalize strategic objectives.

In their various articles and books (Kaplan and Norton 1992, 1996a, b, 2001, 2004a, b, 2006), Kaplan and Norton maintain that use of the BSC concept will ultimately improve an organization's performance. They discuss a variety of ways in which the BSC concept contributes to attaining such improvement. They argue that the BSC "gives top managers a fast but comprehensive view of the business" while minimizing "information overload by limiting the number of measures used" (Kaplan and Norton 1992, p. 72). Additionally, they maintain that the BSC links "a company's long-term strategy with its short-term actions" (Kaplan and Norton 1996b, p. 75). They see this linkage as established through the BSC "translating the strategy to operational terms" and "creating strategic awareness" among employees (Kaplan and Norton 2001, p. 9–11). Strategy can be communicated "to the front lines" (Kaplan and Norton 2001, p. 246), meaning that the "set of hypotheses about cause and effect" (Kaplan and Norton 1996a, p. 149) that constitutes a strategy becomes transparent to all decision makers in an organization. Atkinson (2006) corroborates Kaplan and Norton's arguments. She concludes that the BSC can address a range of common strategy-implementation issues and proposes it as an effective tool against communication deficits, middle-management issues, unclear priorities and lack of coordination, opaque targets, and lacking interpretation of strategic intent into managerial actions.

Fig. 1 Strategy implementation as a first-order control task
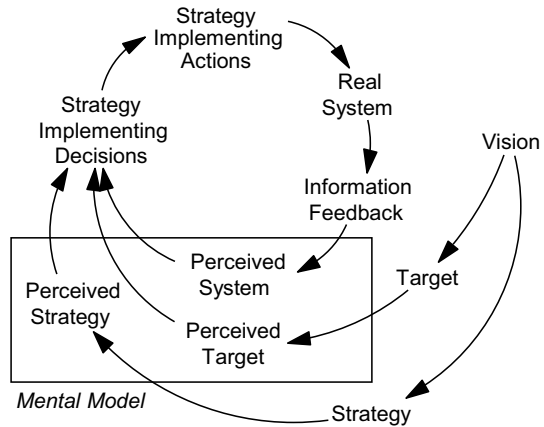
Focusing on the decision-making-related aspects and using the control perspective, as widely suggested in management accounting and operations research, strategy implementation can be seen as a simple, first-order control task (e.g., Dyson 2000; Otley 2003; Sterman 1994). Assuming that (long-term) targets and strategy are derived from a more comprehensive strategic development process, an individual's strategy-implementing decisions are influenced by how the decision maker perceives and processes information about the real system, the targets, and the strategy provided. A gap between the perceived actual and target state of the system would result in decisions guided by the strategy that are translated into actions to close the gap (Fig. 1). According to dynamic decision-making theory, good decision making in dynamically complex settings requires that the "whole process of action regulation" (Dörner and Schaub 1994, p. 434) is conducted successfully: (1) goal elaboration, (2) hypothesis formation, (3) prognosing, (4) planning, (5) monitoring, and (6) self-reflection. An appropriate mental model comprising a clear system of targets, an adequate set of causal hypotheses about the system's causal structure, and an unambiguous understanding of the strategy improves forecasting of the potential actions' consequences, balancing pros and cons and making a choice. It also supports monitoring the results of previous actions and learning from one's own past mistakes. According to experimental research, more accurate mental models do indeed result in better decision-making performance (Capelo and Dias 2009; Gary and Wood 2011; Ritchie-Dunham and Puente 2008).

Using a BSC cockpit has the potential to improve a decision maker's mental model in at least three ways. First, by presenting measures that are directly related to the its strategic themes, strategy is made explicit and operational. Decision makers always have the means by which abstract themes are concretized before their eyes—for instance, the fact that the strategic theme "increase operational excellence" is measured in sales revenue per full-time equivalent employee. Second, by restricting the number of indicators included in a BSC cockpit to 20–25, decision makers' information load is limited. While still above the magical number seven plus or minus two information chunks that human beings can process simultaneously (Baddeley 1994; Miller 1956),

decision makers' information load is nevertheless reduced when compared to the multiple indicators presented in traditional reports. At the same time, a balanced mixture of indicators from four different perspectives minimizes the risk of mistakenly ignoring unintended consequences of previous decisions because information feedback is too focused (Sterman 1994). Third, the four BSC perspectives—and with them their strategic themes and associated measures—are linked by the following logic: improvements in learning and growth measures will be followed by advancements in internal process key performance indicators (KPIs), which will result in better customer perspective indicators and finally translate into higher financial success. Embedding the strategy's strategic themes and measurers in this overall cause-and-effect chain clarifies the causal relations between them and improves decision makers' mental models of the strategy to be implemented. This reasoning leads to the following hypothesis:

H$_{1a}$ Using a BSC cockpit (that is carefully linked to strategy designed according to principles and examples described in the BSC literature) versus a traditional report cockpit by individual managers in their strategy implementation decision-making process is positively associated with strategy implementation performance.

On the other hand, relying on a BSC cockpit in strategy implementation could also come with dangers. Decision makers without adequate knowledge of the BSC concept may find the cockpit design unusual and alien; therefore, they may be not able to make use of the structural information that the cockpit provides. They may not realize the implicit cause-and-effect relationships between the performance indicators. Therefore, the reduction of measures when compared to traditional reports might be harmful, even though cognitive load is reduced. As a consequence, an alternative hypothesis is stated as follows:

H$_{1b}$ Using a BSC cockpit (that is carefully linked to strategy designed according to principles and examples described in the BSC literature) versus a traditional report cockpit by individual managers in their strategy implementation decision-making process is negatively associated with strategy implementation performance.

Psychological research indicates that an individual's performance in dynamic decision-making tasks (such as the strategy implementation task investigated in this study) is also influenced by personal traits—most importantly, intelligence and knowledge (Ackerman 1996). General cognitive ability along with lower-order factors are among the most influential factors that positively relate to dynamic decision-making performance—as long as they are measured carefully (Beckmann and Guthke 1995; Wittmann and Hattrup 2004). Wittmann and Hattrup (2004) emphasize that especially the intelligence sub-factor "reasoning with numbers" is an important predictor of dynamic decision-making performance. Implementing a strategy in a business context naturally requires dealing with numbers and developing a thorough understanding of relations between them. Having a higher intelligence allows participants to build better mental models of the (numerical) targets that they are given, the (numerical) feedback that they receive, and the (verbally and numerically described) strategy that guides their decision making. It is therefore hypothesized that:

H$_2$ The higher decision makers' general cognitive ability, the better they perform in strategy implementation.

Knowledge has long been recognized as having a positive impact on managerial performance, and as differentiating experts and novices (Wagner and Sternberg 1985). Additionally, Krems (1995) argues that knowledge in a domain increases cognitive flexibility in problem solving in that domain and, thus, leads to higher performance in decision-making tasks. Wittmann and Hattrup (2004) find that computer-game-related knowledge is the strongest predictor of performance in the game. Moreover, general economic knowledge also has a positive impact. Therefore, it is reasonable to assume that participants with higher levels of domain- and task-related knowledge can build better mental models, which leads to better decision-making performance. Specifically, if decision makers know more about how to interpret financial and non-financial indicators, they will be better at analyzing the current state of the system and have a more accurate mental representation. Therefore, the following hypothesis is stated:

H_3 The higher the decision makers' (general and performance-measurement-related) knowledge, the better they perform in strategy implementation.

Whether gender makes a difference in dynamically complex decision making is investigated and discussed in various behavioral research streams (Dörner 1996; Eckel and Grossman 2008; Gallagher and Kaufman 2005; Wittmann and Hattrup 2004); the outcomes are mixed, but suggest that the impact of gender on decision-making performance should at least be controlled.

3 Research methodology

Data for testing the causal hypotheses introduced in the previous section could, in principle, be gathered from case studies, surveys, and field and laboratory experiments. The research methodology literature (Cooper and Schindler 2011; Trochim and Donnelly 2008; Zikmund 2012) ranks experimental field studies first in terms of internal and external validity. Therefore, this would be the method of choice in this research, which aims to establish cause-and-effect relationships. However, there are several obstacles that prevent using an experimental field study. Since field studies involve experiments in natural settings, random events and complexity are the most severe issues. In order to isolate the causal relationships between an organization's usage of a BSC cockpit and its strategy implementation success, other factors also impacting performance have to be controlled. However, organizations face randomness and a complex network of cause-and-effect relationships that affect their strategy implementation performance. It is very difficult—if not impossible—to keep track of all those possibly disruptive factors (Sprinkle and Williamson 2007, p. 416). Therefore, a true and valid field experiment would either be extremely costly or irreproducible.

In contrast to most other empirical BSC studies mentioned above, which use the field approach or quasi-experimental designs, this study used the experimental laboratory method. More precisely, a randomized two-group design was implemented (Trochim and Donnelly 2008). With this design, all conditions are the same for both the experimental groups, with one exception: the first group is exposed to the treatment of a BSC cockpit, while the second is provided with a traditional reports cockpit. Additionally, pre- and post-trial questionnaires are administered to gather information

on the participants' pre-experience and knowledge, as well as on their self-assessment and opinions.

While internal validity of the classic randomized two-group design is high, the external validity of a laboratory experiment is always problematic (Levitt and List 2007a, b). The inevitable artificiality of the laboratory might prevent the results from being honestly generalized. In this study, external validity is improved by designing the experiment carefully and choosing a task that is as close to reality as possible. Therefore, following dynamic decision-making and system dynamics research (e.g., Brehmer and Dörner 1993; Dörner 1996; Paich and Sterman 1993; Wittmann and Hattrup 2004), strategy implementation was designed as a dynamically complex, feedback-rich, and path-dependent decision-making task.

4 Strategy implementation task

Implementation of a strategy has to happen through a series of operational decisions. Although in reality more or less all management levels are involved, top-management decisions play an important role in the implementation process. Therefore, this study focuses on top-management decision making; as a consequence, participants in the experiment had to act as top managers. They were given a virtual 10-year (40-quarter) contract for the position as managing director (CEO) of a recently founded mortgage brokerage business called eHypo. Their main task was to successfully implement eHypo's existing and very ambitious organic growth strategy by repeatedly making decisions on prices and resources. They did not have the option to develop a new strategy themselves. The business concept, long-term targets, strategy, and means of intervention were set by the capital owners. However, as eHypo's CEO, the participants had complete control of the three strategy implementation levers available, which could be adjusted quarter by quarter. These levers included (i) the target commission as price control parameter, (ii) the target number of employees, and (iii) the expenditures for developing the business concept and technology (investment in technology). Quarterly marketing expenditure was determined by a simple decision rule: 2.6 % of forecasted sales revenues were spent on marketing.

eHypo's vision, as defined by the capital owners, was described as follows: "In 10 years we want to significantly increase the wealth of our shareholders and at the same time retain our independence. To achieve this, we want to be the very best of residential lending brokers taking the lead in market share, income, profitability and awareness levels." The vision was operationalized by the following set of strategic goals that the participants should achieve simultaneously by making good implementation decisions. Sales revenue was said to grow from €0.3 to €31 million per quarter within a 10-year time-frame, while maintaining profitability throughout the period and keeping eHypo independent. Return on sales (ROS) was to be greater than or equal to 10 % all (or at least most of) the time, and eHypo's market share in the mortgage brokerage business was to grow to 20 %. EVA—calculated following standard practice[2]—was to reach at least €40 million by the end of year 10. At that point,

---

[2] EVA is determined as profit after tax less a capital charge, the latter being the product of the weighted average cost of capital and the economic capital.
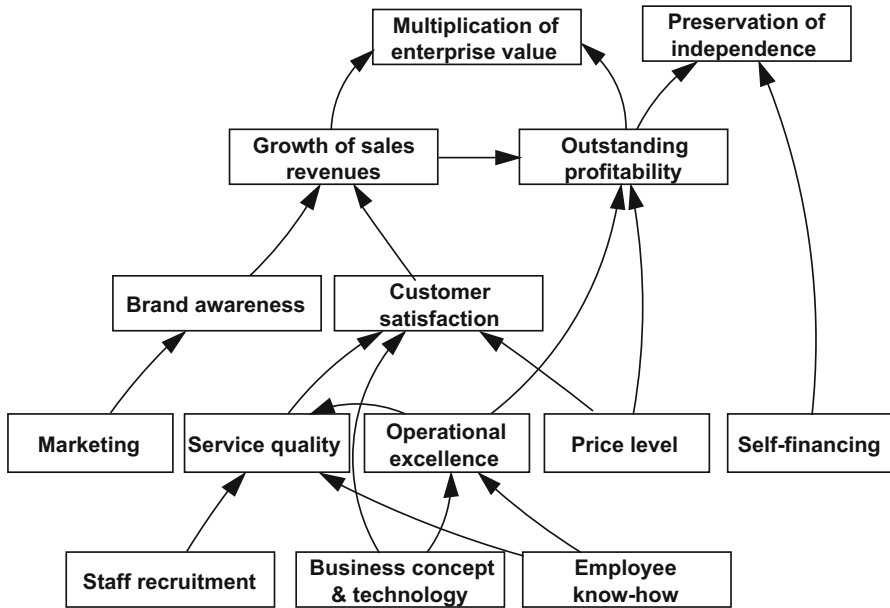
Fig. 2 eHypo's strategy map

eHypo should have been in a solid state, allowing continuation of the business without recapitalization. In addition to this set of strategic goals, an aggregated performance score (P) was provided. The participants were informed that this score was highly positively related with the detailed goals, and that values of 10,000 and higher could be achieved, indicating excellent strategy implementation success. Without revealing the mathematical formula, the instructions described P as the weighted sum of EVA and cumulative absolute ROS variance, which was then multiplied by the attractiveness index that covers the going concern principle. Compared to the EVA measure, P reacts more sensitively to opportunistic end-of-game behavior: for instance, cutting back investments and increasing prices in the last few quarters.

Instructions handed out to the participants included the eHypo strategy paper, which discussed 14 strategic issues that were regarded as important for successfully implementing the growth strategy. It also contained Fig. 2, showing causal links between the 14 strategic issues, providing something close to a strategy map (Kaplan and Norton 2004a). Thereby, all participants had identical and comprehensive information on the strategy to be implemented.

Critical to the growth strategy described in the instructions was the accumulation of eHypo's key resources: staff, technology, employee know-how, and brand awareness (Dierickx and Cool 1989). This growth process should be initiated and maintained without jeopardizing service quality and customer satisfaction in order to avoid the trap of the growth and underinvestment archetype (Senge 1990). eHypo was described as preferring a differentiation strategy over a cost-leadership strategy. High service quality guaranteed by well-trained employees and up-to-date technology were to provide the possibility of escaping sole price competition. Logically consistent with this, the

pricing strategy was set to sustain a medium to high price level, compared to competitive mortgage brokers. Since customers were described as increasingly price sensitive, decreasing prices over time could nevertheless be expected.

In the experimental session, a computerized business game built on system dynamics principles (Sterman 2000) and specifically developed according to the eHypo case was used (Strohhecker and Größler 2012). By design, the instructions and the strategy map reflected the causal relationships incorporated in the game. The game was completely deterministic: two identical simulation game runs—specifically, two identical sequences of decisions over 40 quarters—led to exactly the same outcome. It was also ensured that participants could successfully implement the strategy within the simulator and fulfill or even outperform the ambitious long-term goals set by the owners. Consequently, participants were in a more comfortable situation than real managers. They neither had to react to random events nor deal with the question of whether the strategy itself represented a winning or losing proposition.

Due to the complexity and nonlinearity of the system representing eHypo's business model and environment, an optimal solution cannot be determined analytically. Using a set of common bounded-rational business policies, however, a successful implementation of the strategy set by the owner can be derived by simulation-based policy optimization (Coyle 1985). First, investment in technology was derived from a decision rule that calculated expenses for technology as 4.6 % of forecasted[3] sales. Second, the target number of employees was calculated from forecasted inquiries multiplied by forecasted employee productivity. Lastly, the target value for the commission was based on a policy that started with a reference value of 1.12 %, which was adjusted to market maturity (influencing customers' sensitivity to price) and average capacity utilization. Figure 3 shows the time patterns resulting from these policies.

The participants were able to successfully implement the growth strategy by increasing the target value for employees and the investments in technology in an S-shaped manner, as depicted in Fig. 3. Simulation of the strategy implementation decisions shown in Fig. 3 leads to successful execution of the intended growth strategy in terms of all relevant measures, as Fig. 4 demonstrates. Cumulative EVA would exhibit a hockey-stick shape, which is typical for many start-up organizations. Sales revenue and market share would show a typical S-shaped growth pattern, while ROS could be steadily increased to the target level of about 11 % after an initial period of fast growth and stagnation. The Pearson correlation table in the bottom right of Fig. 4 shows that all measures used to evaluate strategy implementation performance are highly and significantly correlated.

5 Laboratory experiment

To allow thorough reading of the 12 pages of eHypo instructions, the document was handed out to participants one week before the experiment. It was the only input given in advance. The business game in the laboratory was conducted using the eHypo

---

[3] For this and all other forecasts used in decision rules, a simple trend extrapolating forecast function was used (Sterman 2000).
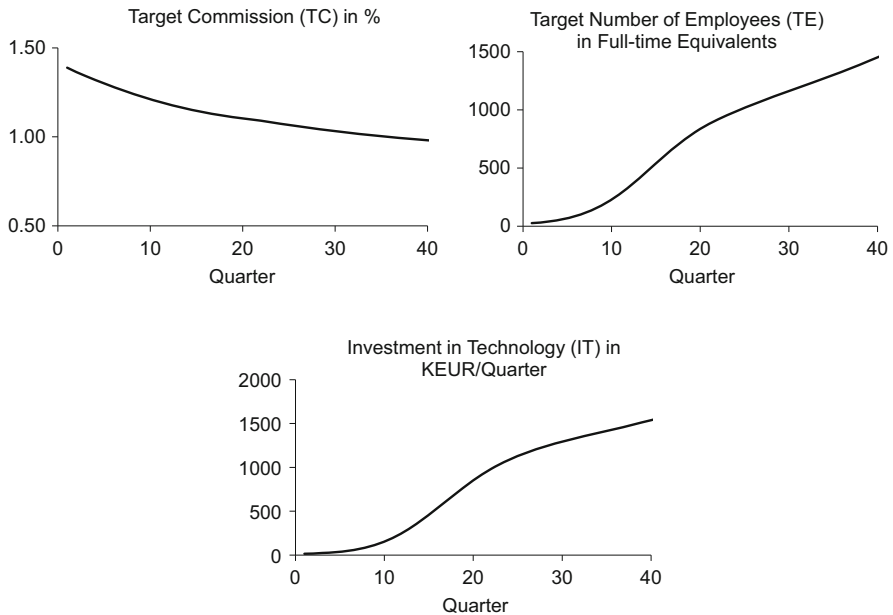
Fig. 3 Decisions leading to the benchmark strategy implementation scenario

simulator software (Strohhecker and Größler 2012). Having made and entered their decisions, participants could continue by clicking a button and simulating one quarter ahead. The outcomes of their decisions were computed, and the updated values for all measures and the overall performance score were displayed. A simulation run included 40 quarters with decisions to be made in each quarter on each of the three decision variables.

The eHypo simulator included the possibility to show the gaming results to the individual participants by two different means: traditional reports (Fig. 5) and a BSC measurement system (Fig. 6). As can be seen, the BSC management cockpit reduces the number of measures displayed compared to the reports cockpit. In accordance with Kaplan and Norton (1996a) recommendation, eHypo's BSC includes 23 measures, while the report cockpit shows 52 indicators. Additionally, the BSC cockpit organized the information in a different way. Figures are linked to eHypo's strategic themes that are themselves related to the BSC's typical four perspectives—learning and growth, internal processes, customers, financials. Embedded in the BSC concept is the idea of cause-and-effect chains stretching over these four perspectives (Kaplan and Norton 1996a, pp. 30–31). Investments in learning and growth strategic measures are assumed to improve internal processes, which positively affects customers. Satisfied customers are supposed to buy more frequently, in larger quantities, and/or to spread word-of-mouth more intensively. All this should result in improved financial measures. The cockpit's design is strongly influenced by Kaplan and Norton (2001, p. 221) proposition for a monthly BSC report. Following this example, the eHypo BSC cockpit organizes the perspectives along this overall causal idea, starting with the learning and growth perspective at the bottom and ending with the financial perspective at the top.

EVA, cumulated (EVAC) in KEUR

Sales Revenue (SR) in KEUR/Quarter

Return on Sales (RoS)

Market Share (MS)

Performance (P) in Points

Pearson Correlations

	1	2	3	4
1 P				
2 EVAC	.983**			
3 SR	.772**	.875**		
4 ROS	.649**	.776**	.981**	
5 MS	.590**	.728**	.966**	.986**

**. Correlation is significant at the 0.01 level (2-tailed).
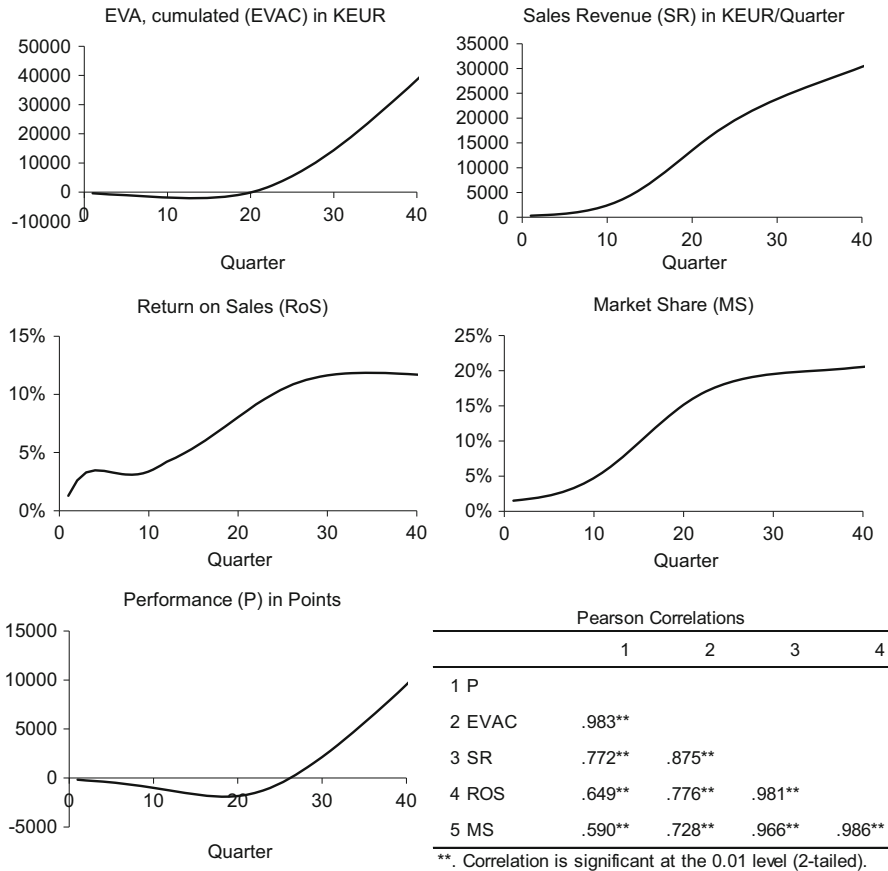
Fig. 4 A successful implementation of the growth strategy used as a benchmark

The report cockpit in Fig. 5 includes income statement and balance sheet as classic ways to structure and communicate information. Six extra reports focus on business development, research and development, cash flows, customer feedback, growth potential, and EVA. As a consequence, the report cockpit provides more detailed information, although this is not specifically related to the strategy that should be implemented.

To exclude the impact of different visualization techniques on performance, which is investigated, for example, by Coll et al. (1994) and Harvey and Bolger (1996), both cockpits show numbers only. Both show the values for the actual and previous quarter for each figure.

In May and June 2010, five experimental sessions were performed, involving a total of 133 participants. Second-semester students enrolled in full- and part-time Bachelor of Business Administration programs at a private German business school were used as participants. All laboratory sessions were integrated in a course on Managerial Accounting, which also covered the BSC concept. The experiment took place in the second to last lecture of the course. Participants were separated according to gender and then assigned randomly to the two different treatments. This ensured that both

Income Statement

Quarter	0	-1
Sales Revenue	162	163
- Labour Costs	96	96
- Marketing Costs	3	3
- Other Operating Expenses	33	34
Administration	16	16
Recruitment	1	1
Redundancy Payments	0	0
Misc	16	16
- Depreciation	66	76
Operating Profit	**-36**	**-45**
- Interest Expenses	49	55
Profit before Tax	**-85**	**-100**
- Taxes	0	0
Profit/Loss	**-85**	**-100**
in 1000 € (K€) per Quarter		

Quarterly Financial Statement

Quarter	0	-1
Assets		
Intangible Assets	361	416
Cash	5.681	6.015
Sum Assets	6.041	6.432
Liabilities		
Equity Prev. Quarter	3.872	3.972
Profit/Loss	-85	-100
Dividends	0	0
	3.787	3.872
Short Term Liabilities	2.254	2.560
Sum Liabilities	6.041	6.432
in 1000 € (K€) per Quarter		

Business Development

Quarter	0	-1
Employee		
Number of Employees	8,0	8,0 Pers.
Employee Expansion	0,2	0,2 Pers./Quarter
Employee Attrition	0,2	0,2 Pers./Quarter
Labour Turnover Rate	2,9	2,9 %/Quarter
Service Capacity	672	678 Enquiries/Quart.
Performance		
Number of Enquiries	1.035	962 Enquiries/Quart.
Number of Contracts Signed	101	102 Deals/Quarter
Brokered Loan Volume	16,18	16,34 Mio. €/Quarter
Market Share		
in the Brokering Segment	1,2	1,4 %
in the Mortgage Loan Market	0,0	0,0 %
Miscellaneous		
Return on Sales	-52,5	-61,4 %
Investment in Marketing		
as % of Sales Revenue	5,4	5,1 %
Price Deviation from Competition	46	47 K€/Pers./Quart.

Research & Development

Quarter	0	-1
Investement in Business		
Concept & Technology (BCT)	10,5	10,5 K€/Quarter
Value of Business Concept		
& Technology	360,9	416,2 K€
Investment in BCT as		
% of Sales Revenue	6,5	6,4 %

Cash Flows

Quarter	0	-1
Cash Flow	5.681	6.015 K€
Operating Cash Flow		
(After Taxes)	-19,1	-24,2 K€/Quarter
Cash Flow from		
Investment Activities	-10,5	-10,5 K€/Quarter
Cash Flow from		
Financing Activities	-305	-347 K€/Quarter

Customers' Feedback

Analysis of customer surveys and Online-Feedback

Quarter	0	-1
Service	62,1	63,3 %
Price Level	100,0	100,0 %
Concept & Technology	99,7	100,0 %

The customers reply to the question about their satisfaction. In case of fulfilling the demands, this corresponds to a satisfaction of 100 %. Percentages with more than 100 percent imply that the customers' wishes are overfulfilled.

Growth Potential

Quarter	0	-1
New Mortgage Loans		
[Bn €/Quarter]	35,82	35,82
Change in New		
Mortgage Loans [%]	0,01	0,01
Brokers' Market Share		
in the Mortgage Loan		
Market [%]	3,9	3,5

EVA

Quarter	0	-1
EVA, cumulated		
[Bn €/Quarter]	-562	-388,8
EVA [K€/Quarter]	-173,2	-186,4
Cost of Capital		
[K€/Quarter]	187,3	187,1
WACC		
[%/Quarter]	4,6	4,6

Fig. 5 eHypo simulator's report-based management cockpit

experimental groups included a similar ratio of women to men. The first group only had access to the BSC report and did not have the traditional form available; the second group was equipped with simulators that only showed the traditional reports. Of course, both groups had the same starting situation in terms of KPI values. But they always saw all the information in their specific cockpit format only.

For both groups, descriptive data on the participants as well as information on their prior knowledge were gathered. Knowledge domains that potentially have to be considered relevant for this study include strategy implementation, performance measurement, the BSC concept, general business knowledge, and computer knowledge. With the exception of performance measurement, knowledge differences are supposed to be negligible. Therefore, only instruments to assess the participants' knowledge and experience in dealing with performance measures were included in the pre-game questionnaire. A second questionnaire was used to gather post-game assessments (see Table 5 in the appendix).

The simulation game was conducted in two different labs, one for each treatment, which eliminated the risk of information exchange between groups. To incentivize the participants, a small impact of the performance in the simulation experiment on the course grade was communicated (e.g., Guala 2005). The best students could achieve up to 5 % of the total points awarded for this course. The higher the aggregate performance measure P in the best simulation run out of the first three in quarter 40 (MXR_P@40), the more points were given.

Measuring general cognitive ability by a standard test such as the Wechsler Adult Intelligence Scale (WAIS–IV) is time consuming. Instead of overloading the pre-trial phase by conducting such a lengthy test, students' entrance assessment-center data

Balanced Scorecard					
Perspective	Strategic Theme	Performance Indicators	Actual Quarter	Previous Quarter	Units of Measurement
1 Finances					
	1 Multiplication of Enterprise Value				
		EVA, cumulated	-389	-202	K€
	2 Preservation of Independence				
		Equity	3982	3983	K€/Quarter
		Short Term Liabilities	78.0	88.0	%/Quarter
	3 Growth of Sales Revenues				
		Sales Revenue	274	242	K€/Quarter
	4 Outstanding Profitability				
		Profit after Tax	-1	-17	K€/Quarter
		Return on Sales	-0.4	-7	%
2 Customers					
	1 Brand Awareness				
		Number of Enquiries	948	904	Enquiries/Quarter
		Brokers' Market Share	1.4	1.4	%
	2 Customer Satisfaction				
		Satisfaction with Price	37.2	37.4	%
		Satisfaction with Service Quality	62.0	61.4	%
		Satisfaction with Technology	99.8	100.0	%
3 Internal Processes					
	1 Marketing				
		Invest. in Marketing as % of Sales Rev.	5.4	5.1	%
	2 Service Quality				
		Service Capacity Utilization	104.9	123.0	%
		Service Capacity	906	736	Enquiries/Quarter
	3 Operational Excellence				
		Sales Revenue per Employee	25.9	29.2	K€/(Empl.* Quart.)
	4 Price Level				
		% Deviation from Competition	47.0	51.8	%
	5 Self-financing				
		Cash Flow	38	35	Enquiries/Quarter
		Cash	3673.0	3635.0	%
4 Learning & Growth					
	1 Business Concept and Technology (BCT)				
		Investment in BCT as % of Sales Revenue	7.5	6.4	%
		Intangible Assets	387	436	K€
	2 Staff Recruitment				
		Number of Employees	14.0	10.2	Persons
		Labour Turnover Rate	2.2	2.8	%/Quarter
	3 Employee Know How				
		Productivity	76	84	Enquiries/Employee

Fig. 6 eHypo simulator's BSC management cockpit

were retrieved from the school's database. This included tests with tasks similar to those from intelligence inventories that aim to measure numerical ability and reasoning (G_NR).[4] Test results on a scale from 1 to 10 could be obtained. In addition, the Abitur grade (AG) was used as a proxy for cognitive ability. Admittedly, these data stem from archival sources and are a few years older than the data gathered in the laboratory. However, psychological research shows that general cognitive ability is rather stable over long time periods (Larsen et al. 2008; Lyons et al. 2009), which seems to justify the use of these data.

---

[4] The following task serves as an example of the type of problems the applicants had to solve in the admission test: A plus B equals 9. B plus B equals 8. A = ?, B = ?. (Solutions: A = 5, B = 4.)

Participants' knowledge about performance measures was assessed by two scales from the pre-game questionnaire. First, the participants were asked to provide a self-assessment of their experience in dealing with financial measures (K_KPI_SA) on a Likert scale of ten items (0 = no experience at all, 9 = highly experienced). Second, specific knowledge about financial KPIs was assessed using a set of eight exam-like questions that the participants had to answer before the treatment (K_KPI).[5] Gender (MALE), age (measured in years, AGE_Y), and participation in an integrated degree program (PTS), which were used as control variables, could be easily retrieved from the school's records, as participants used their student ID number as identification.

Participants in both groups were provided a generous time-frame of 180 min, allocated as follows: about 30 min for the introduction and for questions and answers; about 30 min for the pre-treatment task; about 105 min for the simulation game; and about 15 min for the post-treatment questionnaire. Within the game, participants could become insolvent and therefore fail. In this case, they would virtually be laid off (the simulation was stopped). However, more than one run covering a maximum of 40 quarters each was allowed. While participants were allowed to restart the simulation as often as they wanted, they were instructed that only the results from the first three runs would be evaluated and included in the incentive scheme. This seemed to be a reasonably high number of simulations to avoid failures, which were solely or mainly attributable to faulty operation of the simulator software. On the other hand, the maximum of three valid simulations narrowed the risk that video game syndrome would distort data ascertainment (Sterman 2006). The number of simulation runs and the duration of each simulation were recorded together with all other results in the simulation data file. On completion of the time allowed, the data files with the simulation results were collected so that the relevant data could be extracted.

6 Presentation of results

Of the 133 participants, 40 (30.1 %) were female and 93 (69.9 %) male. A total of 95 participants studied part-time and were employed by a company for at least 50 % of the time (PTS = 1); 38 participants were full-time students. Furthermore, 125 participants were able to complete at least one simulation run successfully (that is, without going bankrupt); 97 participants (72.9 %) were able to completely avoid bankruptcy (number of insolvencies, NI = 0); and 22 participants fell bankrupt once (including one participant who conducted only one run). Seven participants went bankrupt twice and three times, respectively (14 participants total). Table 1 provides summary statistics on all variables included in the study. Post-trial questionnaire items are presented in Table 5 in the Appendix.

Decision-making behavior in the simulation—in terms of time spent per simulation and number of simulation runs conducted—varied among the participants. Mean time

---

[5] First, the participants were given balance sheet and income statement data for two years and asked to calculate five measures from these data: cash flow, operating cash flow, return on equity, equity ratio, and operative result (EBIT). Second, provided with information on the cost–income ratio and sales revenues for two companies, they were asked to complete statements by using $>$, $<$, or $=$ as, for example: Efficiency of company A is ($>$, $<$, $=$) than efficiency of company B.

Table 1 Descriptive statistics on all variables included in the study

	N	Minimum	Maximum	M	SD	Skewness		Kurtosis	
						Statistic	SE	Statistic	SE
MXR_P@40	125	−6,847.62	6,673.95	128.25	1,886.83	0.173	0.217	2.354	0.430
EVAC@MXR_P	125	−8,143.67	39,019.46	8,197.33	10,280.66	0.405	0.217	−0.536	0.430
SR@MXR_P	125	144.82	37,656.86	13,443.41	10,494.71	0.392	0.217	−1.009	0.430
MS@MXR_P	125	0.00070	0.25294	0.08140	0.06730	0.540	0.217	−0.785	0.430
ROS@MXR_P	125	−1.2860	0.20326	0.07921	0.14017	−7.844	0.217	73.998	0.430
MALE	133	0	1	0.70	0.460	−0.879	0.210	−1.246	0.417
AGE_Y	132	18.8	25.6	21.1	1.376	1.110	0.211	0.921	0.419
PTS	133	0	1	0.71	0.453	−0.960	0.210	−1.096	0.417
NR	133	1.00	19.00	4.0451	2.191	3.552	0.210	17.973	0.417
NR_TT3	133	1	3	2.95	0.242	−5.939	0.210	38.390	0.417
NI	133	0	3	0.43	0.819	1.996	0.210	3.170	0.417
MR_TPS	125	5.97	47.33	22.94	8.666	0.551	0.217	0.021	0.430
TPS@MXR_P	125	3.80	70.76	20.86	12.569	1.667	0.217	3.053	0.430
G_NR	123	0.20	0.90	0.58	0.187	0.236	0.218	−0.731	0.433
G_NR_I	133	0.20	0.90	0.59	0.181	0.238	0.210	−0.604	0.417
AG	124	1.00	3.50	2.03	0.489	0.236	0.217	0.189	0.431
AG_I	133	1.00	3.50	2.04	0.479	0.226	0.210	0.251	0.417
K_KPI_SA	132	0	7	2.87	1.826	0.247	0.211	−0.635	0.419
K_KPI	132	0	6	2.74	1.706	−0.208	0.211	−0.914	0.419
PO_P_1	132	2	9	6.95	1.400	−1.261	0.211	2.207	0.419
PO_S_11	132	0	9	7.26	2.310	−1.519	0.211	1.633	0.419
PO_I_2	132	2	9	5.77	1.806	−0.215	0.211	−0.908	0.419
PO_I_5	132	0	9	5.97	1.934	−0.793	0.211	0.243	0.419

Table 1 continued

	N	Minimum	Maximum	M	SD	Skewness		Kurtosis	
						Statistic	SE	Statistic	SE
PO_I_6	132	1	9	6.76	1.708	−0.781	0.211	0.215	0.419
PO_I_7	132	0	9	6.33	1.847	−0.792	0.211	0.633	0.419
PO_I_8	130	0	9	5.71	1.815	−0.266	0.212	−0.301	0.422
PO_I_8_1	133	0	9	5.70	1.797	−0.262	0.210	−0.256	0.417
PO_I_9	132	0	9	6.09	2.174	−0.730	0.211	−0.263	0.419
PO_I_10	132	0	9	5.92	2.162	−0.858	0.211	0.382	0.419
PO_I_15	132	2	9	6.98	1.565	−0.859	0.211	0.692	0.419
PO_I_16	132	2	9	6.17	1.604	−0.390	0.211	−0.253	0.419
PO_I_17	132	0	9	5.48	2.473	−0.283	0.211	−0.675	0.419
PO_DMB_12	132	2	9	6.51	1.422	−0.905	0.211	0.770	0.419
PO_M_13	132	0	9	8.12	1.198	−2.941	0.211	15.674	0.419
PO_M_141	132	0	9	7.04	2.127	−1.471	0.211	2.184	0.419
PO_M_142	132	3	9	7.27	1.462	−1.125	0.211	0.912	0.419
PO_M_143	132	1	9	7.30	1.591	−1.491	0.211	2.566	0.419
PO_U_4	132	0	8	5.21	1.985	−0.567	0.211	−0.244	0.419
PO_A_3	132	0	8	3.93	1.997	−0.027	0.211	−0.846	0.419
Valid N (listwise)	112								

spent per simulation (MR_TPS) ranged between 5.97 and 47.33 min (mean: 22.94 minutes). The time spent on the run with the highest performance (TPS@MXR_P) varied even more: between 3.80 and 70.76 min. Both variables are significantly negatively correlated with the number of simulation runs (NR) conducted (see also Table 2). The majority of participants (57.1 %) ran the eHypo simulator exactly three times. Very few completed fewer than three runs. 39.1 % of the participants ran (voluntarily) more simulations than the three that were incentivized. As a consequence, the number of simulation runs were truncated to a maximum value of three (NR_TT3). Both TPS@MXR_P and NR_TT3 are included as control variables in the analysis.

Figure 7 provides an overview of the results the participants achieved in their runs, with the highest performance P at the end of quarter 40 (MXR_P@40) out of the first three runs. Comparing these outcomes with the benchmark shown in Fig. 4, it is obvious that no single participant was able to implement the growth strategy as successfully as a set of relatively simple heuristics. The highest-performing candidate achieved a performance score MXR_P@40 of 6,673.95. Mean performance was 128.25. Considering EVA totaled over all 40 quarters [cumulated economic value added (EVAC)], 90 out of 125 (72 %) non-bankrupt participants managed to create value. A total of 28 % of all participants destroyed company value, meaning they achieved a negative EVAC at best. On average, 8,197.33 KEUR was accumulated. One reason for this underachievement is that the ambitious growth target set by the owners was rarely met. While the benchmark implementation resulted in a sales revenue of 30,972.92 KEUR in quarter 40 and a market share of 20.68 %, participants achieved on average 13,443.41 KEUR sales revenue and 8.14 % market share. Another reason can be found in the ROS measure: 73.6 % of the participants achieved less than the benchmark of 11.65 % in quarter 40. That the participants were on average less successful in implementing the growth strategy is not a matter of concern here. Given the research question, it is the degree of variation in strategy implementation performance that is of interest, not its shortfall in comparison to the benchmark.

To test hypotheses H_{1a} and H_{1b} (that a BSC cockpit increases/decreases performance) and hypotheses H_2 and H_3 (that intelligence and knowledge increase performance), BSC usage is dummy coded and the following straightforward regression is estimated (controlling for the number of simulation runs NR_TT3, the number of insolvencies NI, the time spent in the simulation run that resulted in the maximum aggregate performance TPS@MXR_P, part-time studies PTS, age AGE_Y, and gender MALE):

Model A:
$$MXR\_P@40 = \beta_0 + \beta_1 BSC\_Cockpit + \beta_2 G\_NR + \beta_3 AG + \beta_4 K\_KPI\_SA + \beta_5 K\_KPI + \beta_6 NR\_TT3 + \beta_7 NI + \beta_8\ TPS@MXR\_P + \beta_9 PTS + \beta_{10} AGE\_Y + \beta_{11} MALE + \varepsilon$$

As a first step in the statistical analysis, both the data generated in the laboratory experiments and those obtained from databases were carefully screened following the guidelines provided by Tabachnick and Fidell (2013). K_KPI_SA, K_KPI and AGE_Y were found to be missing for one case, which was excluded from further analysis. Additionally, G_NR was found to be missing for 10 cases, AG for nine cases, and PO_I_8 for two cases. Listwise inclusion of cases in the regression or correlation analysis reduced the sample size to 113 or 112, respectively, which might

Table 2 Pearson correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1 MXR_P@40																																			
2 EVAC@MXR_P	.773**																																		
3 SR@MXR_P	.400**	.794**																																	
4 MS@MXR_P	.406**	.780**	.995**																																
5 ROS@MXR_P	.282**	.069	-.172	-.141																															
6 MALE	.167	.368**	.334**	.317**	-.112																														
7 AGE_Y	-.016	.053	.110	.110	-.030	.169																													
8 PTS	.129	.212*	.164	.159	.055	.052	.069																												
9 NR	-.141	-.139	-.087	-.097	.055	.146	-.117	-.079																											
10 NR_TT3	-.314**	-.261**	-.164	-.167	-.039	-.110	-.060	.079	.171																										
11 NI	-.373**	-.197*	-.020	-.020	-.163	.014	-.022	-.010	.036	.089																									
12 MR_TPS	.151	.372**	.435**	.427**	-.188*	.199*	.099	.057	-.292**	-.428**	.085																								
13 TPS@MXR_P	.175	.290**	.298**	.296**	-.031	.100	.001	.002	-.287**	-.145	.083	.544**																							
14 G_NR_I	.187	.245*	.123	.111	-.006	.269**	-.104	.173	.030	-.009	.040	.048	-.055																						
15 AG_I	.062	-.038	-.048	-.035	.014	.077	.220*	.194*	.196*	.114	-.207*	-.201*	-.261**	-.052																					
16 K_KPI_SA	.223*	.181*	.120	.134	-.020	.275**	.204*	-.166	.083	-.081	-.150	.044	.004	-.076	.150																				
17 K_KPI	.281**	.351**	.268**	.255**	.064	.299**	.204*	.045	-.039	-.182*	.025	.156	-.015	.140	-.119	.178																			
18 PO_P_1	.217*	.241**	.155	.147	.011	.088	.047	-.054	.075	-.163	-.168	.174	.013	.043	.119	.216*	.142																		
19 PO_S_11	-.092	-.098	-.086	-.098	-.003	-.119	-.019	.045	-.006	.387**	.013	-.358**	-.135	-.052	.051	-.005	-.086	-.115																	
20 PO_I_2	.341**	.285**	.154	.161	.037	.118	.056	-.154	-.238**	-.093	-.121	.305**	.347**	.055	.054	.209*	-.003	.341**	-.105																
21 PO_I_5	.122	.093	.023	.030	.028	.086	.099	.100	-.118	-.163	.104	.137	.006	.014	.160	.077	-.084	.439**		.289**															
22 PO_I_6	.147	.181*	.177	.174	.060	.204*	-.024	-.077	.032	-.012	.145	.201*	-.104	.126	.011	.170	-.115	.175	-.036	.305**	.289**														
23 PO_I_7	.103	.032	.062	.062	.128	-.117	-.150	.023	.141	.029	.078	-.178	-.104	.016	-.023	-.111	.071	-.009	-.007	.008	.014	.080													
24 PO_I_8_I	.282**	.262**	.212*	.215*	.045	.090	.089	-.127	-.192*	-.131	-.059	.216*	.250**	-.036	.009	.065	.014	.157	-.179*	.476**	.351**	.303**	.064												
25 PO_I_9	-.032	-.068	-.082	-.082	-.048	-.065	.016	-.336**	-.019	-.076	-.142	.148	.017	-.005	.044	.122	-.135	.186*	-.122	.308**	.484**	.270**	-.040	.214*											
26 PO_I_10	.012	.018	-.028	-.032	-.009	.078	-.075	-.277**	-.006	-.139	-.043	.225*	.080	.060	.021	.145	-.062	.168	-.168	.261**	.430**	.424**	-.041	.307**	.647**										
27 PO_I_15	-.041	.058	.104	.108	-.117	.168	.003	-.040	-.083	-.029	.035	.201*	.115	-.088	-.091	.084	.024	.138	-.033	.252**	.154	.195*	-.045	.019	.050	.135									
28 PO_I_16	-.126	.003	.125	.126	-.145	.090	.130	.041	-.082	-.035	.092	.206*	.182*	-.130	.010	-.087	-.020	-.143	-.057	-.112	-.118	-.161	-.125	-.136	-.156	-.070	.258**								
29 PO_I_17	.093	.178*	.238**	.249**	.004	.020	.040	.041	-.126	-.089	.002	.189*	.106	-.138	-.004	.077	.031	.130	-.075	.169	.220*	.125	.175	.255**	.040	.185*	.151	.161							
30 PO_DMB_12	.311**	.305**	.191*	.179*	.264**	.095	.065	-.048	.025	-.260**	-.215*	.328**	.182*	.010	.095	.232*	.186*	.353**	-.237**	.299**	.119	.126	-.117	.209*	.098	.224*	.214*	.096	.029						
31 PO_M_13	.076	.066	.030	.024	-.096	.095	.025	-.046	-.019	-.094	-.140	.135	.107	-.183*	.043	.224*	.043	.356**	-.121	.183*	.052	.083	-.189*	.155	.047	.056	.287**	.029	-.104	.549**					
32 PO_M_141	-.015	-.055	-.045	-.041	-.134	.017	.179*	.096	-.009	-.036	.066	-.101	-.055	-.186*	.160	.014	-.151	-.060	.038	-.086	-.090	-.024	-.099	.090	-.088	-.167	.046	.102	.067	.181*	.252**				
33 PO_M_142	.247**	.282**	.222*	.220*	-.092	.186*	.043	.158	.086	-.154	-.193*	.181*	.052	-.013	.210*	.270**	.095	.288**	-.207	.259**	.195*	.249**	-.001	.156	.207*	.143	.158	-.042	.113	.502**	.427**	.141			
34 PO_M_143	.070	.162	.108	.093	-.061	.197*	.066	-.150	.066	-.022	-.094	.156	.074	-.022	.221*	.159	.099	.299**	-.155	.196*	.078	.124	-.088	.015	.223*	.191*	.209*	.148	.069	.482**	.311**	.069	.539**		
35 PO_U_4	.351**	.388**	.286**	.293**	.114	.197*	.098	-.116	-.049	-.160	-.168	.239**	.116	.070	.172	.110	.014	.210*	-.231*	.449**	.362**	.302**	.040	.504**	.085	.236**	.144	-.026	.234**	.333**	.214*	.095	.228*	.158	
36 PO_A_3	.475**	.460**	.300**	.303**	.076	.159	-.020	-.119	-.017	-.185*	-.237**	.347**	.136	.144	.138	.163	.048	.342**	-.219*	.509**	.294**	.226*	.069	.383**	.252**	.314**	-.045	-.124	.157	.327**	.067	-.161	.273**	.153	.496**

N=122   * Correlation is significant at the 0.05 level (2-tailed).   ** Correlation is significant at the 0.01 level (2-tailed).

Pearson Correlations				
	1	2	3	4
1 MXR_P@40				
2 EVAC@MXR_P	.755**			
3 SR@MXR_P	.419**	.800**		
4 MS@MXR_P	.421**	.784**	.995**	
5 ROS@MXR_P	.412**	.152	.018	.027

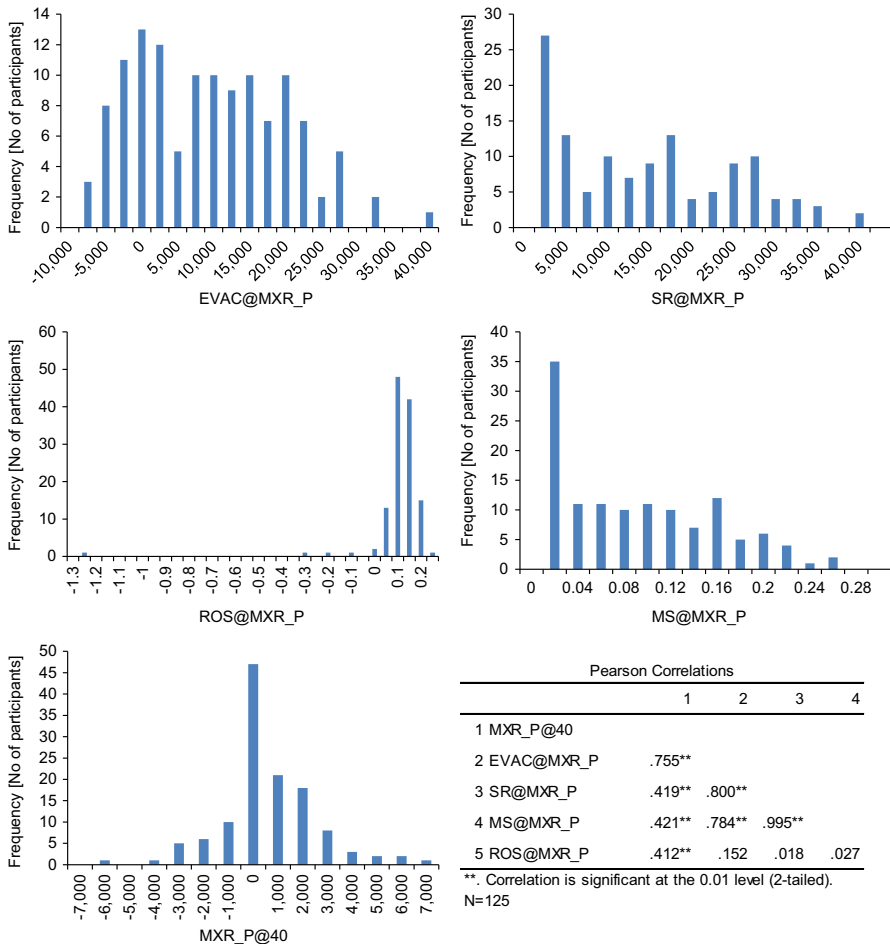**. Correlation is significant at the 0.01 level (2-tailed).
N=125

Fig. 7 Descriptive and correlation statistics on performance scores

have distorted the results. Therefore, missing values in G_NR, AG, PO_I_8 were singly imputed with estimated values (G_NR_I, AG_I, PO_I_8_I) using the expectation-maximization method (e. g., Schafer and Graham 2002). Two cases with standardized residuals greater than 3 were identified and excluded as outliers, leaving an N of 122 for further analysis. A comprehensive set of Pearson correlations is shown in Table 2.

Table 3 shows the regression results for model A (and additional models B–E, explained below). Testing the typical assumptions for regression does not show any violations. Residuals are normally distributed (Kolmogorov–Smirnov test, $Z = 0.686$, $p = 0.734$), and, based on examination of the residuals plot, the assumptions of linearity and homoskedasticity can be considered met. According to the Durbin–Watson test statistics, errors are independent (1.767). Based on maximum variance inflation factors of 1.331, multicollinearity is not considered problematic.

In addition to setting the dependent variable to the best aggregate performance measure P in quarter 40 (MXR_P@40), models B–E use the sub-goals EVAC, sales

Table 3 Regression results for models A–E

Model	A			B			C			D			E		
Dependent Var.	MXR_P@40			EVAC@MXR_P			SR@MXR_P			ROS@MXR_P			MS@MXR_P		
	B	SE B	β	B	SE B	β	B	SE B	β	B	SE B	β	B	SE B	β
(Constant)	6844	3111	0.000**	22,709	18,117	0.000	8641	21,103	0.000	0.167	0.154	0.000	0.060	0.137	0.000
BSC_Cockpit	−42	258	−0.012	−917	1499	−0.045	−1293	1746	−0.062	0.012	0.013	0.086	−0.007	0.011	−0.056
G_NR_I	1484	747	0.159**	8060	4348	0.147*	1753	5064	0.031	0.011	0.037	0.030	0.010	0.033	0.027
AG_I	583	304	0.163*	1907	1772	0.091	1440	2065	0.066	0.000	0.015	0.002	0.011	0.013	0.080
K_KPI_SA	156	76	0.167**	520	443	0.095	220	516	0.039	0.000	0.004	0.005	0.002	0.003	0.059
K_KPI	255	84	0.252***	1582	487	0.266***	1195	567	0.195**	−0.002	0.004	−0.055	0.007	0.004	0.184**
NR_TT3	−2286	746	−0.237***	−8847	4344	−0.156**	−4576	5060	−0.078	−0.021	0.037	−0.055	−0.032	0.033	−0.086
NI	−1034	245	−0.325***	−3525	1428	−0.188***	−449	1663	−0.023	−0.019	0.012	−0.155	−0.002	0.011	−0.017
TPS@MXR_P	31	11	0.226***	243	62	0.303***	243	73	0.294***	0.000	0.001	−0.021	0.002	0.000	0.295***
PTS	659	295	0.176**	4707	1717	0.214***	3695	2000	0.163*	0.009	0.015	0.063	0.024	0.013	0.168*
AGE_Y	−184	101	−0.148*	−610	586	−0.083	−6	682	−0.001	0.000	0.005	−0.008	0.000	0.004	−0.004
MALE	−150	331	−0.039	4017	1929	0.175**	4897	2247	0.206**	−0.017	0.016	−0.110	0.028	0.015	0.186**
R-squared	0.402			0.414			0.255			0.054			0.243		
Adj. R-squared	0.342			0.356			0.181			−0.041			0.167		
N	122			122			122			122			122		

*,**,*** indicates significance at the 90, 95, and 99 % level, respectively

revenues (SR), ROS, and market share (MS) as dependent variables. With the exception of model D where residuals are not normally distributed (Kolmogorov–Smirnov test, $Z = 1.903$, $p = 0.001$), all other typical regression assumptions are met.

Based on the regression results in Table 3, neither H_{1a} nor H_{1b} is supported. In all models, the dummy variable BSC_Cockpit has an insignificant impact on performance. This means that participants using a BSC management cockpit do not make significantly better strategy implementation decisions than participants in the group equipped with a cockpit showing more traditional reports. This finding is supported by the independent sample t-test results compiled in Table 4.

Hypothesis H_2 was operationalized as follows: The higher the decision makers' numerical ability and reasoning, the better they perform in strategy implementation. Regression results from Table 3 support this. In model A, both proxies for cognitive ability, AG_I and G_NR_I, indicate a significant impact on MXR_P@40; in model B, G_NR_I, has a significant impact on the cumulated economic value performance measure (EVAC). Similar results are obtained for the performance measurement knowledge constructs. Both self-assessed knowledge about financial indicators (K_KPI_SA) and knowledge assessed in a test (K_KPI) positively relate to MXR_P@40. K_KPI is also highly significant in models B, C, and E. Therefore, H_3, which states that the higher the decision maker's knowledge the better he or she performs in strategy implementation, can also be seen as supported.

All control variables show a weak significant effect on the dependent variable in at least one model. Regarding the controls that capture certain aspects of the participants' decision-making behavior—the number of runs conducted (NR_TT3) and the time spent for the best run (TPS@MXR_P)—opposite effects can be observed. Higher numbers of simulation runs translate into lower performance in models A and B. More time spent on a simulation run has a highly significant positive effect on performance in all models but D. Similarly, being enrolled in an integrated degree program (PTS), which allows gaining additional experience "on the job," is beneficial in all models but D. Gender (MALE) is significant in models B, C, and E. The age of the participants has a negative effect that is weakly significant only in model A.

7 Discussion

This study shows that two different ways of presenting financial and non-financial performance indicators that support individual participants in implementing an ambitious growth strategy do not make a difference. Organizing these indicators in a decision support cockpit according to BSC principles does not result in higher decision-making performance compared to providing these indicators in a more traditional type of cockpit. However, it does no harm either. The decision-makers' mental models seem insufficiently changed by just using different types of cockpits to make an impact. However, a range of personal factors do show a significant positive impact on strategy implementation performance. As previously found in dynamic decision-making research (Beckmann and Guthke 1995; Brehmer 1992; Wittmann and Hattrup 2004), higher cognitive ability and better knowledge in performance measurement are supportive. Having some work experience is also positively linked to performance in this study's strategy implementation task, while, at the same time, age is disadvantageous,

Table 4 T-test results for all variables included in the study

	Cockpit				t	df
	Reports		BSC			
	M	SD	M	SD		
MXR_P@40	95.28	1770.25	185.45	1699.13	−0.288	120.0
EVAC@MXR_P	8161.92	9916.05	8398.89	10,644.20	−0.128	120.9
SR@MXR_P	13,598.80	9852.78	13,399.02	11,148.11	0.105	120.3
MS@MXR_P	0.08	0.06	0.08	0.07	0.066	120.2
ROS@MXR_P	0.09	0.07	0.09	0.06	−0.746	114.0
MALE	0.70	0.46	0.71	0.46	−0.093	128.1
AGE_Y	21.13	1.42	21.05	1.34	0.366	126.0
PTS	0.71	0.46	0.72	0.45	−0.079	128.1
NR_TT3	2.97	0.18	2.94	0.29	0.645	111.3
NI	0.38	0.79	0.46	0.84	−0.527	128.9
MR_TPS	21.47	8.37	24.34	8.86	−1.844*	121.0
TPS@MXR_P	19.86	13.57	21.69	11.70	−0.800	116.5
G_NR_I	0.60	0.18	0.58	0.18	0.598	128.7
AG_I	2.05	0.46	2.03	0.51	0.265	128.9
K_KPI_SA	2.83	1.74	2.93	1.94	−0.310	127.7
K_KPI	2.54	1.74	2.96	1.67	−1.386	126.7
PO_P_1	6.92	1.36	6.97	1.47	−0.200	128.0
PO_S_11	7.56	2.09	6.93	2.49	1.566	126.3
PO_I_2	5.65	1.88	5.90	1.77	−0.764	126.1
PO_I_5	5.87	2.03	6.10	1.86	−0.677	125.3
PO_I_6	6.68	1.90	6.81	1.52	−0.408	118.7
PO_I_7	6.41	1.79	6.19	1.90	0.675	128.0
PO_I_8_I	5.71	1.71	5.66	1.89	0.174	128.9
PO_I_9	5.75	2.49	6.48	1.78	−1.918*	111.7
PO_I_10	5.71	2.41	6.12	1.91	−1.057	118.1
PO_I_15	7.32	1.42	6.60	1.61	2.711***	127.6
PO_I_16	6.27	1.64	6.04	1.58	0.796	126.8
PO_I_17	5.27	2.44	5.64	2.51	−0.857	127.9
PO_DMB_12	6.54	1.32	6.45	1.53	0.368	127.0
PO_M_13	8.21	0.92	8.03	1.43	0.844	113.6
PO_M_141	7.21	1.82	6.91	2.40	0.793	122.6
PO_M_142	7.32	1.38	7.19	1.55	0.481	127.6
PO_M_143	7.38	1.56	7.22	1.64	0.560	128.0
PO_U_4	5.08	1.93	5.37	1.99	−0.854	127.9
PO_A_3	3.63	2.05	4.24	1.87	−1.752*	125.0

* $p < .1$, ** $p < .05$, *** $p < .001$. Equal variances not assumed

although only weakly significant. As in the Wittmann and Hattrup (2004) study, gender makes a difference in strategy implementation too—at least in performance measures that are directly related to the size of the company: economic value added, sales revenue, and market share. Male participants do a better job at growing the business and therefore achieve better results in these three measures. Regarding the factors that control, to some extent, for dynamic decision-making behavior, time spent per simulation and number of runs conducted, a positive influence of the first and a negative influence of the second is found. Taking more time in a dynamic decision-making task allows for deeper reasoning and seems beneficial in constructing better mental models, which translate into better decisions.

Analyzing the post-trial questionnaire provides additional insights into potential reasons for the failure of the BSC cockpit in improving strategy implementation decision making. Table 5 lists all 19 questions and organizes them in seven categories. The second highest correlation to performance can be seen in item PO_U_4, which measures the clarity of the causal links between decisions and performance criteria. Participants who understand the cause-and-effect chains, which relate their decisions to performance indicators, perform better than others who lack this understanding. This finding underlines the theoretical argument on the mental models' important role in decision making. Using the (Kaplan and Norton 1996a, p. 149) definition of strategy as a "set of hypothesis about cause and effect," item PO_U_4 could also be seen as a measure of strategic clarity: a clear understanding of the causal mechanisms that relate decisions to outcomes results in a clear understanding of strategy. Item PO_P_1 addresses strategic clarity from a slightly different angle. The question was: How clear was the strategy described for eHypo AG in the case study? Participants rating higher on PO_P_1 or PO_U_4 also achieved significantly higher performance ($\rho = 0.217$, p = 0.017; $\rho = 0.351$, p = 0.000). These findings support Kaplan and Norton's (e.g., 2004b, p. 100) argument that strategic clarity is an important success factor. Unfortunately, the BSC cockpit used in this study does not contribute to evoking significant differences in clarity of cause-and-effect chains (see t-test results for items PO_P_1 and PO_U_4 in Table 4). At least in the design chosen in this study (following suggestions in the literature), the BSC cockpit did not support participants' understanding of the causal relations between their decisions and consequences. Their mental model did not (or did not sufficiently) change.

Of all post-game question items in the information category, only two turned out to relate significantly positively to strategy implementation performance: strategy reflection of information (PO_I_2) and information adequacy (PO_I_8_I). However, as shown in Table 4, testing for differences between the two groups of BSC cockpit and report users revealed no significant results. In participants' perceptions, arranging measures according to four logically linked BSC perspectives and according to strategic themes does not increase the strategic relationship of the information provided in the cockpit (compared to traditional reports). Moreover, participants do not perceive the information provided through a BSC cockpit as more adequate to control the company than that provided through the report cockpit. Whether they needed more training in the BSC concept to adequately understand the structure behind the information presented in the cockpit, or whether the cockpit itself needed improvement, cannot be answered by this study and must be left for further research.

Significant differences between BSC cockpit and report cockpit users could be observed on three post-trial questions (see Table 4): First, the BSC cockpit provided a more clearly designed and less confusing information screen (PO_I_9). Second, information provided through the report cockpit was considered more multifaceted than through the BSC cockpit (PO_I_15). This last aspect is not really surprising, as the BSC cockpit, by design, showed fewer measures than the report cockpit, and the report cockpit showed a broad range of non-financial indicators. Third, participants with a BSC cockpit rated implementing the strategy with available information as easier than participants with a report cockpit (PO_A_3). While the post-trial items PO_I_9 and PO_I_15 are not significantly correlated with participants' strategy implementation performance, PO_A_3 shows the highest correlation of all post-trial questions with MXR_P@40. Overall, this effect does not result in a performance difference between the two groups; nevertheless, it is remarkable that participants with access to a BSC cockpit perceive their task as easier than the report cockpit group. This might contribute to explaining positive evaluations of the BSC concept in survey based research (e.g., De Geuser et al. 2009; Rigby 2001; Speckbacher et al. 2003) that does not always and necessarily translate into increases in financial performance (e.g., Ittner et al. 2003).

8 Conclusions, limitations and implications for further research

This study contributes to the BSC and performance measurement literature (Atkinson 2006; Biggart et al. 2010; Davis and Albright 2004; De Geuser et al. 2009; Iselin et al. 2008; Kaplan and Norton 2004b; Tapinos et al. 2011) by showing that a dynamic strategy implementation task is not improved when a BSC cockpit is provided instead of a report cockpit. Participants in a laboratory experiment are not able to make better decisions when faced with a strategy-related reduced set of indicators that are grouped into the four classic BSC perspectives: learning and growth, internal processes, customers, and financials. Participants in the group who are forced to use reports such as balance sheets and income statements do not perform worse. Interestingly, they rate it as less easy to implement the strategy with the available information than the BSC cockpit group. This finding has important implications for BSC designers and adopters because it highlights that success in strategy implementation cannot be achieved just by changing the design of a management information system cockpit. While a BSC cockpit was preferred as less confusing and more clearly designed over a report cockpit by the participants, this did not translate to better decision making. Investing significant amounts of money into developing a BSC cockpit might increase subjective user satisfaction when working with these tools, but might not contribute to increasing more objective performance indicators, supporting similar findings by Ittner et al. (2003). However, it has to be highlighted that this study's results should not be falsely generalized to discourage applying and implementing the BSC approach as a whole, as the concept is much more comprehensive than the focus of this research.

Contributing to dynamic decision-making research (Brehmer 1992; Dörner et al. 1994; Moxnes 1998; Sterman 1989; Wittmann and Hattrup 2004), this study argues that strategy implementation can be seen a typical dynamic decision-making task. While obviously being less complex than developing and implementing a strategy, executing a specified growth strategy nevertheless poses a huge challenge for human

decision makers. Findings on the human "logic of failure" (Dörner 1996) in dynamic decision-making research are supported by this study's results. In addition, further evidence is provided that cognitive ability and knowledge have a significant positive impact on decision-making performance—not only in the classic complex problem-solving environment (Brehmer 1992; Brehmer and Dörner 1993), but also in a more operational setting, supporting findings by Strohhecker and Größler (2013). In line with recent research on the relation between time on task and ability in complex problem solving (Scherer et al. 2015), this study highlights that taking more time in dynamic decision making is beneficial. Deep thinking, understood as reasoning with numbers (Wittmann and Hattrup 2004), which also in this study shows a positive impact on decision making outcomes, should therefore be complemented by long thinking.

The results contribute to the strategy literature (Ansoff 1984; Dyson 2000; Kaplan and Norton 2008; Mintzberg 1990) by showing that bad execution is not only an organizational phenomenon, but can also be rooted to some extent in decision-maker characteristics, as mentioned above. From this individual perspective, creating strategic clarity seems an important success factor. Participants who had a clearer understanding of the causal links between their actions and the results achieved thereby performed significantly better. However, the BSC cockpit used in this study failed to increase strategic clarity perceived by the participants (compared to the report cockpit). With its tabular, dashboard-like design that closely follows suggestions by Kaplan and Norton, cause-and-effect chains are not visualized and therefore not emphasized. The BSC's main causal chain—better measures in the learning and growth perspective lead to improved indicators in the internal processes perspective that result in better customer performance indicators and improved financial measures—might be too invisible and perhaps also too generic to be helpful for strategy implementation decision makers. Whether improved cockpit designs that clarify the strategy as a "set of hypotheses about cause and effect" (Kaplan and Norton 1996a, p. 149), or whether better-educated and more experienced users who could get more information out of the BSC cockpit would change this study's findings, has to be left for further research.

Of course, this study is not without limitations. When focusing on strategy implementation, giving participants the task to decide on execution only seems justified. However, most CEOs have more competences: they do not only implement an already specified strategy; they can also develop one. Therefore, future research could give the participants more "power" and define the task more broadly. The focus on strategy implementation results in a second limitation. Participants need to have a very thorough description of the strategy that they have to implement, including a strategy map that outlines the important cause-and-effect relationships between strategic themes. Giving this information in very detailed form to both experimental groups might have reduced the discriminating effect of a BSC cockpit. Future research could investigate whether different forms of strategy descriptions, such as verbal only versus verbal and graphical, resulted in significant performance differences. Another limitation comes from using a realistic, rather complex case study and business game in this experiment. Achieving good strategy implementation results may need more repetitions. Therefore, a future study could allow participants more simulation runs that count, and/or could use a less complex strategy implementation case study. Another limitation comes from using a student sample in this research. While more than two-thirds of

participants had work experience, they could not be expected to hold positions where they had strategy implementation power to the extent reflected in the experiment. They might also not have the type of experience that allows for a different decision-making style; for instance, more intuitive decision making. Therefore, future research could use participants from the population of more experienced managers and investigate whether the factors affecting strategy implementation performance change.

9 Appendix

Table 5 Post-trial questionnaire and variable coding

Category	Question	Variable label
Preparation	How clearly was the strategy described for eHYPO AG in the case study? (0 = totally unclear; 9 = clear as daylight)	PO_P_1
Situation	How large was time pressure during simulation? (0 = very large; 9 = very small)	PO_S_11
Information	How well did the information provided in the eHYPO simulator reflect the strategy? (0 = extremely poor, 9 = very good)	PO_I_2
Information	How do you assess the quality of the presentation of the information in the eHYPO simulator? (0 = extremely poor, 9 = very good)	PO_I_5
Information	How comprehensible was the information provided? (0 = incomprehensible; 9 = fully understandable)	PO_I_6
Information	How well did you cope with the quantitiy of information provided? (0 = too few information, 9 = too much information)	PO_I_7
Information	Was the information adequate for you to control the company? (0 = not at all; 9 = entirely appropriate)	PO_I_8
Information	How clearly designed were the available information screens? (0 = unclear and confusing; 9 = very well arranged)	PO_I_9
Information	How helpful was the information arrangement? (0 = not at all helpful; 9 = very helpful)	PO_I_10
Information	How multifaceted was the company information? (0 = very loopsided; 9 = very multifaceted)	PO_I_15
Information	How high did you perceive the grade of complexity of the information? (0 = very low; 9 = extremely high)	PO_I_16
Information	How do you assess the extent of the past data provided? (0 = insufficient; 9 = fully adequate)	PO_I_17
Decision making behavior	How well did you think through your decisions? (0 = not at all; 9 = very good)	PO_DMB_12
Motivation	How large was your motivation for achieving a very good result in simulation? (0 = very small; 9 = extremely high)	PO_M_13
Motivation	How strongly did the following elements encourage you? Additional college credits (0 = not at all; 9 = very strong)	PO_M_141
Motivation	How strongly did the following elements encourage you? Fun (0 = not at all; 9 = very strong)	PO_M_142

Table 5 continued

Category	Question	Variable label
Motivation	How strongly did the following elements encourage you? Interest in topic (0 = not at all; 9 = very strong)	PO_M_143
Understanding	How clear were the connections between your decisions and the results achieved thereby for you? (0 = totally unclear; 9 = clear as daylight)	PO_U_4
Achievement	How easily did you implement strategy with available information? (0 = very difficult; 9 = very easy)	PO_A_3

References

Ackerman, P. L. (1996). A theory of adult intellectual development: process, personality, interests, and knowledge. *Intelligence*, *22*, 227–257.

Alexander, J. (2007). *Performance dashboards and analysis for value creation*. Wiley, Hoboken, NJ: Wiley finance.

Alexander, L. D. (1985). Successfully implementing strategic decisions. *Long Range Planning*, *18*, 91–97.

Ansoff, I. (1984). *Implanting strategic management*. London, Englewood Cliffs, N. J: Prentice-Hall.

Atkinson, H. (2006). Strategy implementation: a role for the balanced scorecard? *Management Decision*, *44*, 1441–1460.

Baddeley, A. (1994). The magical number seven: Still magic after all these years? *Psychological Review*, *101*, 353–356. doi:10.1037/0033-295X.101.2.353.

Beckmann, J. F., & Guthke, J. (1995). Complex problem solving, intelligence, and learning ability. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 177–200). Hillsdale: Erlbaum.

Biggart, T. B., Burney, L. L., Flanagan, R., & Harden, J. W. (2010). Is a balanced scorecard useful in a competitive retail environment? *Management Accounting Quarterly*, *12*, 1–12.

Brehmer, B. (1992). Dynamic decision making: human control of complex systems. *Acta Psychologica*, *81*, 211–241.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, *9*, 171–184. doi:10.1016/0747-5632(93)90005-d.

Bryant, L., Jones, D. A., & Widener, S. K. (2004). Managing Value Creation within the Firm: An Examination of Multiple Performance Measures. *Journal of Management Accounting Research*, *16*, 107–131.

Bungay, S., & Goold, M. (1991). Creating a strategic control system. *Long Range Planning*, *24*, 32–39. doi:10.1016/0024-6301(91)90182-n.

Cândido, C. J. F., & Santos, S. P. (2015). Strategy implementation: What is the failure rate? *Journal of Management & Organization*, *21*, 237–262. doi:10.1017/jmo.2014.77.

Capelo, C., & Dias, J. F. (2009). A system dynamics-based simulation experiment for testing mental model and performance effects of using the balanced scorecard. *System Dynamics Review*, *25*, 1–34. doi:10.1002/sdr.413.

Coll, R. A., Coll, J. H., & Thakur, G. (1994). Graphs and Tables: A Four-Factor Experiment. *Communications of the ACM*, *37*, 77–86.

Cooper, D. R., & Schindler, P. S. (2011). Business research methods. In The McGraw-Hill/Irwin series in operations and decision sciences, 11th edn. New York: McGraw-Hill/Irwin

Coyle, R. G. (1985). The use of optimization methods for policy design in a system dynamics model. *System Dynamics Review*, *1*, 81–91. doi:10.1002/sdr.4260010107.

Daum, J. H. (2006). Management Cockpit War Room: Ziele, Funktionsweise und Zukunftsperspektiven eines (noch) ungewöhnlichen, aber hocheffektiven Managementinstruments. *Controlling*, *18*, 311–318.

Davis, S., & Albright, T. L. (2004). An Investigation of the Effect of Balanced Scorecard Implementation on Financial Performance. *Management Accounting Research*, *15*, 145–153.

De Geuser, F., Mooraj, S., & Oyon, D. (2009). Does the Balanced Scorecard Add Value? Empirical Evidence on its Effect on Performance. *European Accounting Review*, *18*, 93–122. doi:10.1080/09638180802481698.

Dierickx, I., & Cool, K. (1989). Asset Stock Accumulation and Sustainability of Competitive Advantage. *Management Science*, *35*, 1504–1511.

Dörner, D. (1980). On the difficulties people have in dealing with complexity. *Simulations and Games*, *11*, 87–106.

Dörner, D. (1996). *The logic of failure*., Strategic thinking for complex situations New York, NY: Metropolitan Books.

Dörner, D., Kreuzig, H. W., & Reither, F. (1994). *Lohhausen*. Huber, Bern et al: Vom Umgang mit Unbestimmtheit und Komplexität.

Dörner, D., & Schaub, H. (1994). Errors in Planning and Decision-making and the Nature of Human Information Processing. *Applied Psychology*, *43*, 433–453. doi:10.1111/j.1464-0597.1994.tb00839.x.

Dyson, R. G. (2000). Strategy, performance and operational research. *Journal of the Operational Research Society*, *51*, 5–11.

Eckel, C. C., & Grossman, P. J. (2008). Differences in the Economic Decisions of Men and Women: Experimental Evidence. In: R. P. Charles, L. S. Vernon (Eds.), Handbook of Experimental Economics Results, vol 1 (pp. 509–519). Amsterdam: Elsevier. doi:10.1016/S1574-0722(07)00057-1.

Eckerson, W. W. (2011). *Performance dashboards: Measuring, monitoring, and managing your business* (2nd ed.). Hoboken, NJ: Wiley.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors*, *4*, 59–73.

Gallagher, A. M., & Kaufman, J. C. (2005). *Gender differences in mathematics: an integrative psychological approach*. Cambridge: Cambridge University Press.

Gary, M. S., & Wood, R. E. (2011). Mental models, decision rules, and performance heterogeneity. *Strategic Management Journal*, *32*, 569–594. doi:10.1002/smj.899.

Guala, F. (2005). *The methodology of experimental economics*. Cambridge: Cambridge University Press.

Harvey, N., & Bolger, F. (1996). Graphs versus tables: effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, *12*, 119–137.

Hoque, Z., & James, W. (2000). Linking balanced scorecard measures to size and market factors: impact on organizational performance. *Journal of Management Accounting Research*, *12*, 1–17.

Iselin, E. R., Mia, L., & Sands, J. (2008). The effects of the balanced scorecard on performance: the impact of the alignment of the strategic goals and performance reporting. *Journal of General Management*, *33*, 71–85.

Ittner, C. D., & Larcker, D. F. (1998). Innovations in performance measurement: trends and research implications. *Journal of Management Accounting Research*, *10*, 205–239.

Ittner, C. D., Larcker, D. F., & Randall, T. (2003). Performance implications of strategic performance measurement in financial services firms. *Accounting, Organizations and Society*, *28*, 715–741.

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard—measures that drive performance. *Harvard Business Review*, *70*, 71–79.

Kaplan, R. S., & Norton, D. P. (1996a). *The balanced scorecard: translating strategy into action*. Boston, MA: Harvard Business School Press.

Kaplan, R. S., & Norton, D. P. (1996b). Using the balanced scorecard as a strategic management system. *Harvard Business Review*, *74*, 75–85.

Kaplan, R. S., & Norton, D. P. (2001). *The strategy-focused organization: how balanced scorecard companies thrive in the new business environment*. Boston, MA: Harvard Business School Press.

Kaplan, R. S., & Norton, D. P. (2004a). How strategy maps frame an organization's objectives. *Financial Executive*, *20*, 40–46.

Kaplan, R. S., & Norton, D. P. (2004b). *Strategy maps: converting intangible assets into tangible outcomes*. Boston, MA: Harvard Business School Press.

Kaplan, R. S., & Norton, D. P. (2006). *Alignment: using the balanced scorecard to create corporate synergies*. Boston, MA: Harvard Business School Press.

Kaplan, R. S., & Norton, D. P. (2008). *The execution premium: linking strategy to operations for competitive advantage*. Boston, MA: Harvard Business School Press.

Krems, J. F. (1995). Cognitive flexibility and complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 201–218). Hillsdale: Erlbaum.

Lane, D. C. (1999). Social theory and system dynamics practice. *European Journal of Operational Research*, *113*, 501–527. doi:10.1016/S0377-2217(98)00192-1.

Larsen, L., Hartmann, P., & Nyborg, H. (2008). The stability of general intelligence from early adulthood to middle-age. *Intelligence*, *36*, 29–34. doi:10.1016/j.intell.2007.01.001.

Levitt, S. D., & List, J. A. (2007). Viewpoint: on the generalizability of lab behaviour to the field. *Canadian Journal of Economics/Revue canadienne d'économique*, *40*, 347–370. doi:10.1111/j.1365-2966.2007.00412.x.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, *21*, 153–174. doi:10.1257/jep.21.2.153.

Lipe, M. G., & Salterio, S. E. (2000). The balanced scorecard: judgmental effects of common and unique performance measures. *Accounting Review*, *75*, 283–299.

Lyons, M. J., et al. (2009). Genes determine stability and the environment determines change in cognitive ability during 35 years of adulthood. *Psychological Science*, *20*, 1146–1152. doi:10.1111/j.1467-9280.2009.02425.x.

Maiga, A. S., & Jacobs, F. A. (2003). Balanced scorecard, activity-based costing and company performance: an empirical analysis. *Journal of Managerial Issues*, *15*, 283–302.

Malina, M. A., & Selto, F. H. (2001). Communicating and controlling strategy: an empirical study of the effectiveness of the balanced scorecard. *Journal of Management Accounting Research*, *13*, 47–90.

Mankins, M. C., & Steele, R. (2005). Turning great strategy into great performance. *Harvard Business Review*, *83*, 64–72.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97. doi:10.1037/h0043158.

Mintzberg, H. (1990). The design school: reconsidering the basic premises of strategic management. *Strategic Management Journal*, *11*, 171–171.

Moxnes, E. (1998). Not only the tragedy of the commons: misperceptions of bioeconomics. *Management Science*, *44*, 1234–1248.

Otley, D. (2003). Management control and performance management: whence and whither? *The British Accounting Review*, *35*, 309–326. doi:10.1016/j.bar.2003.08.002.

Paich, M., & Sterman, J. D. (1993). Boom, bust, and failures to learn in experimental markets. *Management Science*, *39*, 1439–1458.

Person, R. (2008). *Balanced scorecards & operational dashboards with Microsoft Excel*. New York, NY: Wiley.

Reichert, U., & Dörner, D. (1988). Heurismen beim Umgang mit einem "einfachen" dynamischen System. *Sprache und Kognition*, *7*, 12–24.

Rigby, D. (2001). Management tools and techniques: a survey. *California Management Review*, *43*, 139–160.

Ritchie-Dunham, J. L., & Puente, L. M. (2008). Strategic clarity: actions for identifying and correcting gaps in mental models. *Long Range Planning*, *41*, 509–529. doi:10.1016/j.lrp.2008.06.006.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods*, *7*, 147–177. doi:10.1037//1082-989X.7.2.147.

Scherer, R., Greiff, S., & Hautamäki, J. (2015). Exploring the relation between time on task and ability in complex problem solving. *Intelligence*, *48*, 37–50. doi:10.1016/j.intell.2014.10.003.

Senge, P. M. (1990). *The fifth discipline: the art and practice of the learning organization*. New York, NY: Currency Doubleday.

Sim, K. L., & Koh, H. C. (2001). Balanced scorecard: a rising trend in strategic performance measurement. *Measuring Business Excellence*, *5*, 18–28.

Speckbacher, G., Bischof, J., & Pfeiffer, T. (2003). A descriptive analysis on the implementation of balanced scorecards in German-speaking countries. *Management Accounting Research*, *14*, 361–389.

Sprinkle, G. B., & Williamson, M. G. (2007). Experimental research in managerial accounting. In C. S. Chapman, A. G. Hopwood, & M. D. Shields (Eds.), *Handbook of management accounting research* (Vol. 1, pp. 415–444). Amsterdam: Elsevier.

Sterling, J. (2003). Translating strategy into effective implementation: dispelling the myths and highlighting what works. *Strategy & Leadership*, *31*, 27.

Sterman, J. D. (1989). Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. *Management Science*, *35*, 321–339.

Sterman, J. D. (1994). Learning in and about c omplex systems. *System Dynamics Review*, *10*, 291–330. doi:10.1002/sdr.4260100214.

Sterman, J. D. (2000). *Business dynamics: systems thinking and modeling for a complex world*. Boston, MA: Irwin/McGraw-Hill.

Sterman, J. D. (2006). Learning from evidence in a complex world. *American Journal of Public Health*, *96*, 505–514. doi:10.2105/ajph.2005.066043.

Strohhecker, J., & Größler, A. (2012). Implementing sustainable business strategies. *Systems Research and Behavioral Science*, *29*, 547–570. doi:10.1002/sres.2139.

Strohhecker, J., & Größler, A. (2013). Do personal traits influence inventory management performance? The case of intelligence, personality, interest and knowledge. *International Journal of Production Economics*, *142*, 37–50. doi:10.1016/j.ijpe.2012.08.005.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Tapinos, E., Dyson, R. G., & Meadows, M. (2011). Does the balanced scorecard make a difference to the strategy development process? *Journal of the Operational Research Society*, *62*, 888–899. doi:10.1057/jors.2010.99.

Tayler, W. B. (2010). The balanced scorecard as a strategy-evaluation tool: the effects of implementation involvement and a causal-chain focus. *The Accounting Review*, *85*, 1095–1117. doi:10.2308/accr.2010.85.3.1095.

Trochim, W. M. K., & Donnelly, J. P. (2008). *Research methods knowledge base* (3rd ed.). Cincinnati, OH: Atomic Dog Pub.

Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: the role of tacit knowledge. *Journal of Personality and Social Psychology*, *49*, 436–458. doi:10.1037/0022-3514.49.2.436.

Wittmann, W. W., & Hattrup, K. (2004). The relationship between performance in dynamic systems and intelligence. *Systems Research and Behavioral Science*, *21*, 393–409.

Zikmund, W. G. (2012). *Business research methods* (9th ed.). Mason, OH: Erin Joyner.