Do Planning Support Systems Improve Planning? Testing the claim in a controlled experiment

Marco te Brömmelstroet[1], Peter Pelzer[2], Ralph Klerkx[3] and Sander Schaminée[3]

[1] University of Amsterdam, Amsterdam The Netherlands
[2] Utrecht University, Utrecht, The Netherlands
[3] TNO, Utrecht, The Netherlands
email corresponding author: brommelstroet@uva.nl

Introduction

In planning research, it is considered a truism urban planning is about linking different types of knowledge to actions in the public domain (Friedmann 1987). There are many debates about what types of knowledge are needed when (Khakee, Barbanente et al. 2000) and how they should be related to planning actions (Healey 2007, Nonaka, Konno 1998, Forester 1989), but the general added value is widely accepted.

Researchers and practitioners in the field of Planning Support Systems (PSS) start from this notion and have developed a wide range of dedicated instruments to support this process of linking and integrating planning-relevant knowledge, with a special focus on adding scientific and explicit knowledge to specific parts of the planning process (Geertman, Stillwell 2009, Geertman, Stillwell 2003, Brail 2008, Brail, Klosterman 2001).

These instruments hold a huge potential for the "hopelessly complex human endeavor" of urban planning (Couclelis 2005, p. 1355). However, as has been found in a growing body of studies, their use in planning practices is very limited. In his seminal study on this "implementation gap", Vonk sums up these research findings as "[PSS are] far too generic, too complex, too inflexible, incompatible with most planning tasks and oriented towards technology rather than problems and too focused on strict rationality" (Vonk 2006, p. 19). One of the main conclusions of his research

into the underlying reasons for this is that there is a fundamental and consistent gap between PSS developers and their potential users. This mirrors findings of other researchers in this field (e.g. Lee 1973, Lee 1994, authors).

In this light, it is striking that there is hardly a concerted research effort to overcome this implementation gap. A recent study showed that research into the usability of PSS is almost non-existent and fragmented at best (authors). It seems ample time for a coordinated effort to move forward on improving PSS usability (authors). To better understand what effective interventions for this are, we follow a pragmatic research agenda. Based on the concept of 'realistic evaluation' we aim to create insights on what currently (does not) work in which contexts and why (not) (Pawson, Tilley 1997).

The goal of this paper is to develop and use an analytical framework that support this realistic evaluation. First, we define the concept of planning quality into a multi-dimensional framework (section 2). Then we operationalize this framework in section 3, and use it in an experimental setup to test the general added value of a PSS for a typical planning process (section 4). After distilling the relevant conclusions from this, the paper closes with a reflection on the research and a discussion on what it means for PSS research, development and use.

Performance of PSS

In a recent meta-analysis of PSS literature, a conceptual framework was introduced that formulated 'improving planning quality' as core goal of PSS (authors. Based on literature on knowledge management, process management and group model building, this was then translated into thirteen dimensions (Dean, Hender et al. 2006, Rouwette, Vennix et al. 2002). Below, we will shortly address the logic followed.

Planning quality is first divided into quality of outcomes and quality of planning processes. There is no clear academic consensus about what defines a *good* or *bad* planning outcomes. Following widely accepted and used insights from the field of ideational output, we choose to define the quality of a planning outcome into four main dimensions; novelty, workability, relevance and specificity (Dean, Hender et al. 2006). These four dimensions are further broken down into sub dimensions; e.g. novelty into originality and paradigm relatedness; and workability into implementability and acceptability (Table 1).

Table 1. Multi-dimensional framework to measure quality of planning and performance of PSS

Planning outcome		Planning process	
Novelty		Reaction	
Original	A	Enthusiasm	J
Paradigm relatedness	B	Satisfaction	K
Workability		Credibility	L
Implementability	C	Insight	
Acceptability	D	Insight in problem	M
Relevance		Insight in assumptions	N
Applicability	E	Commitment	O
Effectiveness	F	Behaviour	P
Specificity		Communication	Q
Completeness	G	Developm. of shared language	R
Implicational explicitness	H	Consensus	
Clarity	I	Consensus on problem	S
		Consensus on goals	T
		Consensus on strategies	U
		Cohesion	V
		Efficiency gains	W

Quality of planning processes is also a topic that is widely debated in planning literature, Here, we follow the line that a planning process is better if it supported individual and group learning. Note that we acknowledge that this aspect is more important in the more strategic phases of planning, where problems are often 'wicked' (Rittel, Webber 1984). This is also where we use the framework in this research. This concept of learning links to the idea that by adding knowledge (through a PSS), planning participants are better able to extend their personal knowledge with that of the PSS itself and with the personal knowledge of other participants around the table ((authors, Gudmundsson 2011, Amara, Ouimet et al. 2004). Based on Group Model Building research, that specifically focused on supporting (group) learning with instruments (Rouwette, Vennix et al.

2002, Rouwette 2003, Rouwette, Vennix et al. 2009), we defined quality of planning processes into nine dimensions this leads to eight dimensions. Note that the first four dimensions relate to personal learning, whereas the latter five express the quality of the group process. Again, some dimensions are broken down into sub dimensions (table 1).

Operationalisation

Table 2. Statements per dimension for quality of planning outcome

A1	The strategy is ingenious.
A2	The strategy is imaginative.
A3	The strategy is surprising.
A4	The strategy is novel
B1	The strategy is radical.
B2	The strategy is transformational.
C1	The strategy can be easily implemented.
D1	The strategy is socially acceptable.
D2	The strategy is legally acceptable.
D3	The strategy is politically acceptable.
E1	The strategy clearly applies to the stated problem.
F1	The strategy will solve the problem.
F2	This is an effective strategy
G1	The strategy can be decomposed into independent subcomponents.
G2	The strategy covers who
G3	The strategy covers what
G4	The strategy covers where
G5	The strategy covers when
G6	The strategy covers why
G7	The strategy covers how
H1	There is a clear relationship between actions and expected outcome.
I1	The strategy is clearly communicated
I2	The strategy is easy to understand

The use the conceptual framework of planning quality, we need to operationalize it into measureable indicators. To do so, all dimensions are translated into (several) statements (table 2). We stayed close to the statements used in the field of ideational output. These statements are then used to ask external raters to assess the quality of planning outcomes (using a 7-point Likert scale from ‘totally disagree’ to ‘totally agree’).

Second, we have described each subdimension of the quality of the planning process into several statements (table 3). Here, we stayed as close as possible to statements used in Group Model Building research. Participants of planning sessions are asked to respond to these statements (using a 7-point Likert scale from ‘ totally disagree’ to ‘totally agree’).

This operationalisation is sensitive to a wide variety of impacts of a PSS on planning quality. It allows us to test what attributes of planning outcomes and processes benefit from PSS support and can thus be a valuable analytical framework for realistic evaluation of PSS usability.

Research design

Our aim is that the operationalized framework can be used in a wide variety of research settings to perform realistic evaluation of PSS usability and to develop insights on the effectiveness of interventions to improve this. Here, we will use it to test the general hypothesis that PSS have an added value for planning.

Why experiment?

As discussed elsewhere, empirical research on the use of PSS is almost completely based on (often) single case studies (authors); a PSS is designed and applied in a certain planning context. This has provided us with a large body of very rich insights in context specific and situated applications of these instruments. However, this dominant research design poses a severe limitation to the development of a general understanding of PSS usability. The context-embedded nature of most empirical research makes it hard, if not impossible, to isolate effects of a PSS intervention from effects of context variables. This limits the internal and external validity of this research and poses serious limits to establish reciprocal links with the strong theoretical work that has been done (Geertman, Stillwell 2009, Geertman, Stillwell 2003, Brail 2008, Brail, Klosterman 2001).

Table 3. Statements per dimension for quality of planning process

J1	I have a good feeling about the session
J2	The session resulted in valuable results
K1	The session was successful.
K2	I am satisfied with this session.
K3	The other participants are satisfied with the session
L1	The results of the session offer real solutions for the problem
L2	Results of the session are based on correct assumptions on the underlying system
L3	I am confident that the group solution is correct
M1	My insight into the problem has increased.
M2	The session has given me insight into relations of elements that compose the problem.
M3	It is clear to me what the causes of the problem are.
M4	I now have more insight into the processes that play a role in the problem.
M5	The session resulted in new insights.
N1	My understanding of opinions of the other participants about the problem has increased.
N2	I understand how other participants in the session perceive the problem.
N3	Other participants understand how I perceive the problem
N4	I better understand the proposed solutions of other participants in the session.
O1	I support most of the results that were drawn during the session.
P1	I will use insights from the session in my daily planning practice
Q1	The process has given me insight into other people's opinions and ideas about the problem.
R1	During the sessions we have developed a shared professional language
R2	During the sessions a platform emerged that supported the sharing of ideas
S1	We have reached a shared vision of the problem.
S2	The results integrated diverse opinions and ideas of the participants.
S3	We were able to reach a consensus on the problem.
T1	We have reached a shared vision on the strategic goals
U1	We have reached a shared vision on the possible solutions
V1	I had a strong sense of being part of a group
V2	The session brought me closer to the other participants
V3	We experienced conflict during the session
V4	There was conflict about the task we had in the session
W1	The session was time efficient

One research design that supports realistic evaluation and strengthens this link between theory development and empirical research is the controlled experiment; "if a causal factor, X, is manipulated, then, given appropriate controls, a systematic effect is produced on the response variable, Y" (Goldthorpe 2001). If so, this systematic effect is associated with (and probably caused by) the manipulation of the causal factor. Ideas about causal factors are taken directly from PSS theory and outcomes can be translated directly back into this theoretical body. An additional benefit is that the research can be easily replicated to control for different contexts.

Other experiments in PSS literature

There is a small number of studies that applied an experimental design in the PSS literature. Below, we shortly address these.

In a field experiment, Ligtenberg and Vonk compared two PSS development methods as causal mechanism for the acceptance functionality and usability of a sketch PSS (Ligtenberg, Vonk 2010). They used the same group of planners in both control (technical rational development of a PSS) and treatment (socio technical development) conditions. Their case study, as they refer to it, concludes that the socio technical development leads to better results. It could however be argued that treatment and control were not independent and this creates significant noise.

Nyerges et al. tested two different setups of the same PSS (mainly the process of interaction with the instrument) in two different groups (Nyerges, Jankowski et al. 2006) and found effects on the number of options generated, level of consensus and satisfaction of the planners.

In another study Jankowski and Nyerges applied a laboratory experimental setting (Jankowski, Nyerges 2001). 109 voluntary student participants formed 22 groups which received similar support (a PSS with a moderator-chauffeur). Each group had 5 tasks (in random order) that all involved site selection, but varied on complexity, conflict and access to technology. Findings suggest that the maps provided by the technology play a very limited role in the group decision making. There was no control group to which the findings could be compared.

A second laboratory setup is the recent study done by Arciniegas et al. (2013). They set up a laboratory setting with 32 students to test four hypotheses on the effectiveness of three support tools. Each participant used all three PSS to perform the same planning task. Here, the order of use was randomized to control for the independence between the treatments, although they acknowledge that learning could have taken place and could have influenced the results (i.e. perceived and observed effectiveness of

the three tools). There was however no control group, so the study does not give us insights in the general added value of PSS for planning.

Experiment setup

We designed our study as a randomized controlled laboratory experiment (Bryman 2008). With this we aim to optimize the internal and external validity of our findings. Testing the most general claim of the PSS literature (that it improves planning) calls for a strong focus on the ability to translate our findings to theory. Although we have had special attention to mirror characteristics of urban planning practice as good as possible (see below), this means a sacrifice of ecological validity.

Student groups

The experiment was set up as an obligatory part of a second year course of the Bachelor Urban Planning at the University of Amsterdam in November 2012. A total of 78 students participated. They were informed that they took part in a urban planning competition. These students were randomly divided into groups of six. Within each group, each student was, again randomly, assigned one of six planning roles and received information about the plan (see below) that was relevant for his/her specialism and a specific goal for the planning session. Each group consisted of:

- 1 Environmental specialist (air quality)
- 1 Environmental specialist (External safety)
- 1 Environmental specialist (sound)
- 1 Urban Designer
- 1 Transport specialist
- 1 Project economist

Each of the 13 groups got the same planning challenge for a infill area in the old harbors of Rotterdam (figure 1). They were presented with an existing design for the area (figure 2) and the corresponding problems (each role had their own knowledge of specific problematic effects of the plan) and were asked to develop a new plan that would solve these issues as good as possible in a session of 60 minutes.

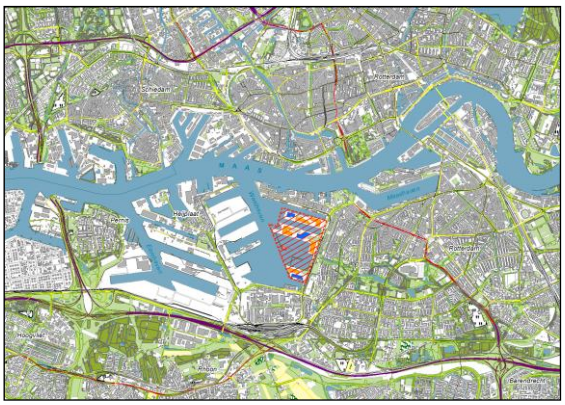

Fig. 1. Infill location in the old harbors of Rotterdam

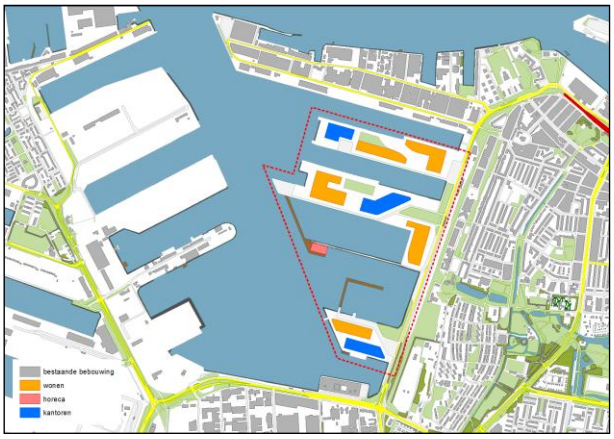

Fig. 2. Orginal design for the area provided to each group

Control versus treatment

The resulting thirteen groups were, again randomly, divided into six control- and seven treatment groups. The control groups received no support; they were assigned to a table with empty plans, instructed to start, and informed that the time was up. These six groups worked simultaneously in one room.

The treatment groups got the full support of a PSS that aims to support these types of planning projects (and actually was used to support planning processes for this area). We have invited this PSS, titled ‘Urban Strategy’, because it represents the state of the art and therefore enables us to test the

general hypothesis in a situation were the added value for planning is mostly expected.

Urban Strategy was developed by TNO and consists of a number of innovative features. First, it links a 2D and 3D design interface with eight different analysis models (transport, noise, air quality, liveability, groundwater, sustainability, external safety and costs modules). Secondly, it can calculate the effects of interventions for all modules within minutes, to enable use in a workshop setting. Thirdly, it uses a surface table to enable participants to interactively engage with the design and effects. Fourthly, and not unimportantly, in this research it offers process support by a team of three people. Two of them mainly served as a chauffeur of the PSS during the treatment. They translated the conversation into interventions in the PSS, ran the models and presented the output back to the group. A third person performed as a process moderator and guided the groups in their design and analysis iterations. These treatment groups worked in series. The physical set up is displayed in figure 3.

Fig. 3. Setup of control and treatment: PSS with surface table, chauffeur and process moderator (left) and the business-as-usual table (right)

Data gathering and analysis

To find out if there are any systematic differences in the performance of the control- and treatment groups we have made use of several data gathering techniques.

First, we have used direct observation by a fourth person: in real time, but also by using video- and audiotaping. Second, all participants filled in an evaluation form in which their personal perceptions of the process quality were asked by responding to the statements of table 2 with a 7-point Likert scale. Third, two external planning experts rated the quality of the resulting strategies (PhD candidates in Urban Planning of the University of

Amsterdam). They were not informed of the hypothesis, nor were they aware of which strategies came from control- or treatment groups.

For the analysis, the responses on the statements were averaged and then compared and tested for systematic differences. To indicate the strength of the differences in effects, we use the p-value of a ANOVA F-test to compare two independent means. For an effect to be considered systematic, it needs to have a p-value of 0,05 or smaller (a conventional statistic measure for significance). The statements were grouped for the subdimensions and overall dimensions by averaging them. Again, the same test was performed. The two statements on conflict were first inverted to make them compatible with this process. For the outcome dimensions, the scores of the raters were also averaged and then processed in the same way.

There were thirteen groups of which seven received our treatment. During the first treatment session, there were severe problems with the process and with the support by Urban Strategy. A quick scan showed that this significantly influenced the results of this group. For the purpose of this research, the results of this group were excluded from further analysis.

Results

In this section, we will present the results on all the dimensions of table 1; first for the outcome- and then for the process dimensions. From that, we will discuss the general trend and the most remarkable findings.

Effects on quality of the outcome

If we look to the overall effect on the quality of the outcome, we compare the quality of twelve different strategies for the Waalhaven area. This resulted in the surprising finding that the PSS treatment resulted in a negative effect on the general quality of the outcome, as operationalized in this research (see table 1 and 2). The score of the control group is already quite low, but the negative difference of the treatment group is remarkable (scale goes from 1 to 7). A possible explanation for this can be the short time (60 minutes) that the groups had to develop their strategy. It is already hard to develop a high quality strategy in this period, understanding and working with the PSS could even put more stress on this task.

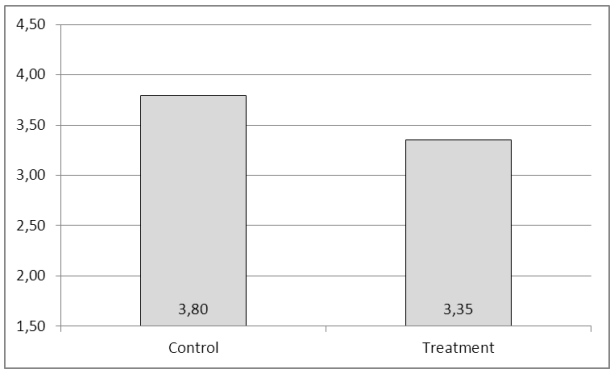

Fig. 4. Overall quality of outcome (p = 0,143)

Table 4. Results planning outcome per dimension

		Average control	Average treatment	Differences	P value
NOVELTY	**3,25**	**2,50**	**-0,75**	**0,101**	
	ORIGINAL	3,06	2,42	-0,64	0,165
	PARADIGMRELATEDNESS	3,63	2,67	-0,96	0,061
WORKABILITY	**4,59**	**4,30**	**-0,29**	**0,455**	
	IMPLEMENTABILITY	4,08	4,25	0,17	0,813
	ACCEPTABILITY	4,75	4,31	-0,44	0,213
RELEVANCE	**3,81**	**3,20**	**-0,61**	**0,199**	
	APPLICABILITY	4,25	3,33	-0,92	0,110
	EFFECTIVENESS	3,58	3,13	-0,46	0,301
SPECIFICITY	**3,83**	**3,53**	**-0,31**	**0,168**	
	COMPLETENESS	3,66	3,32	-0,34	0,123
	IMPLICATIONALEXPLICITNESS	4,33	3,67	-0,67	0,033
	CLARITY	4,21	4,17	-0,04	0,914

Next to this general effect, we can zoom in on the dimension and subdimensions that comprise quality of the outcome (table 4). The second column presents the average of the six strategies of the control groups, the third column the average scores for the treatment groups, the fourth column the difference between the two and the last column the result of the ANOVA t-test. Doing so reveals that the PSS has the biggest negative impact on the novelty of the strategies. Workability is more ambiguous because there is a small positive effect on implementability and a negative

effect on acceptability. Relevance is negatively influenced and here it is especially the applicability of the strategies that scores much lower in the treatment groups. The only systematic effect is found on IMPLICATIONAL EXPLICITNESS (-0,67). It could be that the groups without the PSS felt that they had to be more explicit.

The biggest positive effect of the PSS treatment was found on the statement that addressed the implementability of the strategy (+0,17 (scale 1-7)). It should be noted that this difference if very small and not significant (figure 5). The biggest negative effect was found on the statement that addressed the radicalism of the strategies (-1,08). This effect is also much larger and stronger than the first (figure 5). Also note the big gap in the control group values on both statements: it seems that implementation is already better included than is radicalness. The given goal of the strategy might have had a strong impact on this.

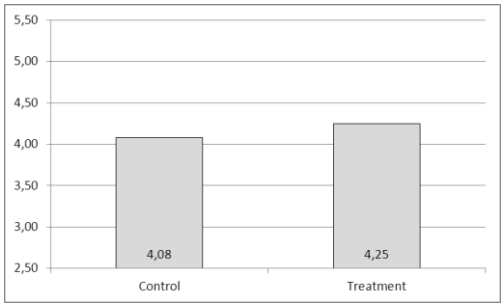

Fig. 5. Average rater response to statement "The strategy can be easily implemented" (p = 0,813)

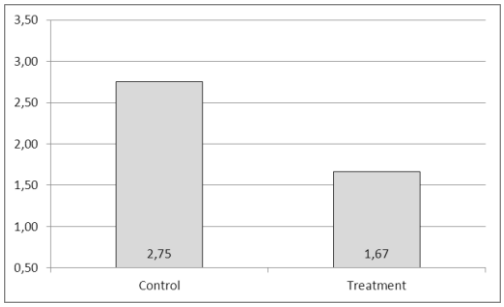

Fig. 6. Average rate response to statement "The strategy is radical" (p = 0,100)

Effects on quality of the process

The comparison of the quality of the process is based on the averaged perceptions of the 84 participating students. In contrast to the quality of the outcome, the quality of the process shows a positive effect of the PSS treatment. Overall, the scores are also higher than on the outcome statements. Still, the treatment has an average that is 0,27 points higher.

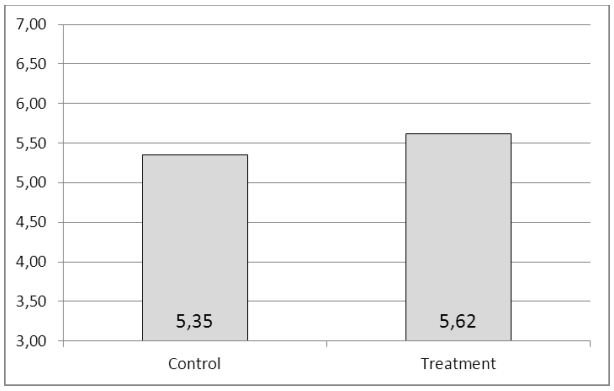

Fig. 7. Overall quality of process (P = 0,100)

Again, we can look at this on the level of the eight dimensions and their sub dimensions. Here we see that there is only one negative effect (COHESION), one without effect (COMMUNICATION) while all other (sub)dimensions show a positive effect. It is remarkable that these remaining positive effects are in the same range (between 0,31 and 0,48). In general, the strength of the positive effects is strong, but systematic effects are only found on REACTION (+0,44), ENTHUSIASM (+0,43), CREDIBILITY (+0,51), CONSENSUS (+0,44), CONSENSUSPROBLEM (+0,44) and CONSENSUSGOALS (+0,45). This indicates that the treatment achieved positive effects on individual learning and on the group process. From observation, the influence of the process moderator and chauffeurs of the PSS seemed to be most relevant on this. Most groups spent considerable time around the design table and used this to share individual ideas before going to the Maptable and calculate effects of these ideas.

In this light, it is surprising to see that there is no measured effect on COMMUNICATION.

Table 5. Results process per dimension

	Average control	Average treatment	Differences	P value
REACTION	**5,38**	**5,82**	**0,44**	**0,035**
ENTHUSIASM	5,45	5,90	0,45	0,043
SATISFACTION	5,61	5,97	0,36	0,068
CREDIBILITY	5,10	5,61	0,51	0,033
INSIGHT	**4,98**	**5,29**	**0,31**	**0,110**
INSIGHTPROBLEM	4,90	5,27	0,38	0,102
INSIGHTASSUMPTIONS	5,07	5,29	0,22	0,298
COMMITMENT	**5,84**	**6,25**	**0,41**	**0,093**
COMMUNICATION	**5,39**	**5,39**	**0,00**	**0,995**
SHAREDLANGUAGE	**4,76**	**5,15**	**0,39**	**0,101**
CONSENSUS	**5,73**	**6,17**	**0,44**	**0,025**
CONSENSUSPROBLEM	5,78	6,24	0,45	0,036
CONSENSUSGOALS	5,58	6,06	0,48	0,030
CONSENSUSSTRATEGIES	5,71	6,09	0,38	0,102
COHESION	**5,47**	**5,32**	**-0,15**	**0,412**
EFFICIENCY	**5,48**	**5,74**	**0,25**	**0,396**

When we look at the individual statements, we find the largest negative effect (-0,41) on the statement that addressed the sense of being part of a group (figure 8). Observations confirmed this: the control groups were challenged to organize themselves, while the treatment groups stayed more passive and followed the process moderator. Both control- and treatment groups score relatively high on this statement. The largest positive effect (+0,75) was measured on the statement that addressed insight into the causes of the problem (figure 9). The PSS was very useful in educating the participants on the underlying dynamics that caused the sectoral problems (on especially air quality and noise) that the groups were confronted with.

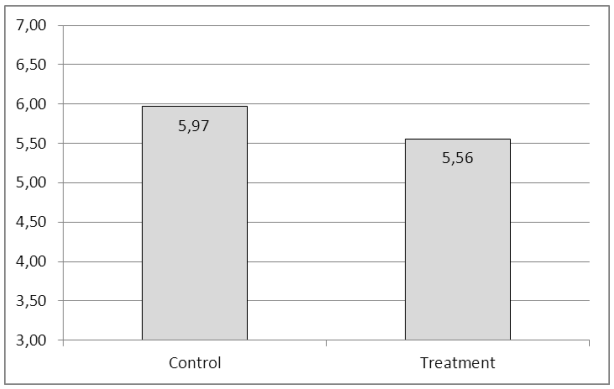

Fig. 8. Average responses to statement 'I had a strong sense of being part of a group' (P = 0,095)

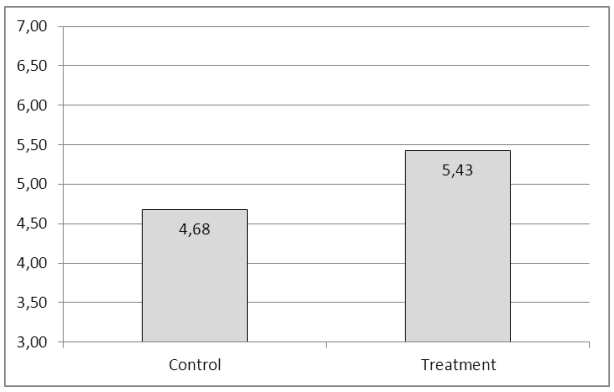

Fig. 9. Average response to statement 'It is clear to me what the causes of the problem are' (P= 0,017)

Conclusions

The setup of the experiment was to test the general premise that PSS have an added value for planning. To test this, we have first defined and operationalized a definition of planning quality. By dividing this concept into the quality of the planning process and the quality of the outcome it was possible to operationalize this concept into measurable statements. The resulting multi dimensional framework was then used analytical tool in an randomized controlled experiment. In this experiment 84 students in 12 groups performed a planning tasks. By controlling the support they re-

ceived during this task (PSS versus 'business as usual') we were able to search for systematic effects on all the dimensions of planning quality.

The findings of the experiment show a mixed image. We mainly used the statements to find if there were effects between the control- and treatment groups. Our own observation was used to find preliminary explanations for these findings.

To answer the general research question: The PSS (Urban Strategy) had an added value for this particular planning situation. It was especially strong in supporting the group process and in providing valuable insights on mechanisms that caused the urban problems that the groups were confronted with. Since the treatment included process and content, it is hard to state what the causes are of these systematic effects. We observed that especially the organization of the process lead to better consensus processes, but (as indicated by the data) to less communication and cohesion in the group.

One of the most remarkable trends in the findings is the large differences between the positive effects on the process quality and the negative effects on the outcome quality. The treatment resulted in (much) lower scores on novelty and relevance of the strategies.

Methodological reflection

We are aware that, although based on relevant academic debates, the definition of planning quality is somewhat subjective and influences the way in which the performance of PSS is measured. By providing the details of how we operationalized this, we aim to be as transparent as possible. As analytical tool in our experiment it reflected our observations and it was sensitive to a wide variety of potential effects of PSS on planning quality. We are open for suggestions on how to further improve the methodological validity of this framework. Special care is needed when the framework is used in more concrete planning phases, where other characteristics of the outcome and the process might be important. Also, we need to test the internal consistency of the statements that aim to measure the same indicator.

Randomized controlled experiments have a number of scientific benefits; maximized internal validity by controlling for the causal variable and high replicability (the procedures are well documented and can be repeated to re-test our findings in the same or in different context and with same or different PSS). However, it severely limits ecological validity. Working with students in a highly controlled laboratory setting limits the transfera-

bility of our findings to real planning practice. Also, external validity is still weak. Do our student groups represent more general features that are important for the use of PSS?

Then there is the question on how close we managed to approach the golden standard of the experimental research design. Findings on causal relations can only be done "given appropriate controls" (Goldthorpe 2001). Although we randomized the students and the groups, the relative small numbers could lead to effects of group composition (only men, only women, mixed, etc.) and of individual characteristics (strong leader type, local knowledge, student performance, etc.). There was one control group that scored much higher on the quality of the planning outcome than the others, which indicates that there some of these effects took place. Role playing was added to make the group dynamics resemble planning practice better, but during the experiment it appeared that most students did not really actively played their role.

Due to budget (of us) and time (of students) limitations, we were only able to observe the treatment groups. The control groups all performed at the same time. Therefore it is not always easy to explain differences in scores.

A final note is that although we measured the added value of a PSS in our experiment, such a research design does only indirectly answer the question of added value for planning practices. To find out more about the relevance of our findings, research in real planning practices is needed.

Discussion

Observations and concrete experiences of the chauffeurs and process mediator immediately led to many improvement ideas for Urban Strategy and the accompanying process. Although we focused on finding an added value of PSS in a controlled environment, many new interesting research question came to the fore.

A first direction for further research is testing interventions that improve PSS usability with the same analytical framework and research design. This allows us to test general ideas that improving transparency, flexibility and communicative value improve PSS usability, and if so which attributes of planning quality profit from his. Our findings indicate that effective improvement interventions should be sought in improving the knowledge exchange process between individual participants and between them and the PSS.

A second research direction is to focus on the intervening variables that influence how the PSS performs or is perceived to perform. The role that a participant played showed some interesting indications of different effects of Urban Strategy: The environmental specialist air quality was much better able to propose interventions from his perspective (5,25 vs. 6,63, p = 0,008) whereas the urban designer was much less able to perform his/her role (5,5 vs. 4,33, p = 0,035). Since we didn't explicitly control for this (and worked with N = 12), more research is needed to establish the validity of these intriguing suggestions. The same can be done for different groups compositions; boy dominated groups are expected to behave differently than girl dominated or mixed groups and groups with a strong leader are also expected to show different characteristics.

Thirdly, the same analytical framework could be used to analyze real world planning settings (with and without PSS) to find out of our findings are replicated there. This would also allow us to improve the measurement validity of the framework itself.

A fourth direction of research could focus on replicating the experiment with other student populations and/or even with participants that are closer to real planning practice. This could strengthen the external validity. One way to move into this direction is to include students from different specialist backgrounds in one experiment.

A final highly interesting research direction is a cross analysis of the findings. With a larger N (after a number of experiments) we should be able to reflect on some general (group-)psychological premises, such as that more comfortable group processes limits the creative ability.

References

ABBOTT, A., 2004. *Methods of Discovery: Heuristics for the social science.* New York/London: W.W. Norton & Company.

AMARA, N., OUIMET, M. and LANDRY, R., 2004. New Evidence on Instrumental, Conceptual, and Symbolic Utilization of University Research in Government Agencies. *Science Communication,* **26**(1), pp. 75-106.

ARCINIEGAS, G., JANSSEN, R. and RIETVELD, P., 2013. Effectiveness of collaborative map-based decision support tools: Results of an experiment. *Environmental Modelling & Software,* **39**(0), pp. 159-175.

BATTY, M., 2003. Planning Support Systems: techniques that are driving planning. In: S. GEERTMAN and J. STILLWELL, eds, *Planning Support Systems in practice.* Heidelberg: Springer, pp. v-viii.

BRAIL, R.K. and KLOSTERMAN, R.E., 2001. *Planning support systems: integrating Geographical Information Systems, models and visualization tools.* Redlands, CA: ESRI Press.

BRAIL, R.K., ed, 2008. *Planning Support Systems for Cities and Regions.* Cambridge, Massachusetts: Lincoln Institute of Land Policy.

BRYMAN, A., 2008. *Social Science Methods.* Third edition edn. Oxford: Oxford University Press.

COUCLELIS, H., 2005. "Where has the future gone?" Rethinking the role of integrated land-use models in spatial planning. *Environment and planning A,* **37**, pp. 1353-1371.

DEAN, D.L., HENDER, J.M., RODGERS, T.L. and SANTANEN, E.L., 2006. Identifying Quality, Novel, and Creative Ideas:Constructs and Scales for Idea Evaluation. *Journal of the Association for Information Systems,* **7**(10), pp. 646-699.

FORESTER, J., 1989. *Planning in the face of power.* Berkeley: University of California Press.

FRIEDMANN, J., 1987. *Planning in the public domain: From knowledge to action.* Princeton: Princeton University Press.

GEERTMAN, S., 2006. Potentials for Planning Support: A planning-conceptual approach. *Environment and planning B : Planning and Design,* **33**(6), pp. 863-880.

GEERTMAN, S. and STILLWELL, J., eds, 2009. *Planning Support Systems: Best Practices and New Methods.* New York: Springer.

GEERTMAN, S. and STILLWELL, J., eds, 2003. *Planning support systems in practice.* Berlin: Springer.

GOLDTHORPE, J.H., 2001. Causation, statistics, and sociology. *European Sociological Review,* **17**(1), pp. 1-20.

GUDMUNDSSON, H., 2011. Analysing Models as a Knowledge Technology in Transport Planning. *Transport Reviews,* **31**(2), pp. 145-159.

HEALEY, P., 2007. *Urban Complexity And Spatial Strategies: Towards A Relational Planning For Our Times.* London: Routledge.

JANKOWSKI, P. and NYERGES, T., 2001. GIS-Supported Collaborative Decision Making: Results of an Experiment. *Annals of the Association of American Geographers,* **91**(1), pp. 48-70.

KHAKEE, A., BARBANENTE, A. and BORRI, D., 2000. Expert and experiential knowledge in planning. *Journal of the Operational Research Society,* **51**(7), pp. 776-788.

KLOSTERMAN, R.E., 2007. Deliberating about the future. In: L.D. HOPKINS and M.A. ZAPATA, eds, *Engaging the future: Forecasts, scenarios, plans and projects.* Cambridge: Lincoln Institute of land policy, pp. 199-219.

LEE, D.B., 1994. Retrospective on large-scale urban models. *Journal of the American Planning Association,* **60**(1), pp. 35-40.

LEE, D.B., 1973. Requiem for large-scale models. *Journal of the American Planning Association,* **39**, pp. pp. 163-178.

LIGTENBERG, A. and VONK, G., 2010. Socio-technical PSS development to improve functionality and usability—Sketch planning using a Maptable. *Landscape and urban Planning,* **94**(3-4), pp. 166-174.

MERTON, R.K., 1973. The Normative Structure of Science. In: R.K. MERTON, ed, *The Sociology of Science: Theoretical and Empirical Investigations.* Chicago: University of Chicago Press, .

NONAKA, I. and KONNO, N., 1998. The concept of 'Ba': Building a foundation for knowledge creation. *California management review,* **40**(3), pp. 40-54.

NYERGES, T., JANKOWSKI, P., TUTHILL, D. and RAMSEY, K., 2006. Collaborative Water Resource Decision Support: Results of a Field Experiment. *Annals of the Association of American Geographers,* **96**(4), pp. 699-725.

PAWSON, R. and TILLEY, N., 1997. *Realistic Evaluation.* London: Sage.

RITTEL, H. and Webber, M., 1984. Dilemmas in a general theory of planning. In:CROSS, N., eds, *Developments in design methodology.* Chicester: John Wiley and Sons , pp. 135-144

ROUWETTE, E.A.J.A., 2003. *Group model building as mutual persuasion.* Nijmegen: Katholieke Universiteit Nijmegen (Ph.D. Dissertation).

ROUWETTE, E.A.J.A., VENNIX, J.A.M. and FELLING, A.J.A., 2009. On Evaluating the Performance of Problem Structuring Methods: An Attempt at Formulating a Conceptual Model. *Group Decision and Negotiation,* **18**, pp. 567-587.

ROUWETTE, E.A.J.A., VENNIX, J.A.M. and VAN MULLEKOM, T., 2002. Group model building effectiveness: a review of assessment studies. *System Dynamics Review,* **18**(1), pp. 5-45.

VONK, G., 2006. *Improving planning support; The use of planning support systems for spatial planning.* Utrecht: Nederlandse Geografische Studies.

VONK, G. and GEERTMAN, S., 2008. Knowledge-Based Planning Using Planning Support Systems: Practice-Oriented Lessons. In: T. YIGITCANLAR, K. VELIBEYOGLU and S. BAUM, eds, *Creative urban regions: harnessing urban technologies to support knowledge city initiatives.* Queensland: Medical Information Science Reference, pp. 203-216.