University of Amsterdam

Centre for Urban Studies

Working Paper Series No. 12

DO PLANNING SUPPORT SYSTEMS IMPROVE PLANNING OR NOT

A measurement framework and randomized experiment

Marco te Brömmelstroet

Centre for Urban Studies
Working Paper
October 2014
www.urbanstudies.uva.nl/workingpapers

Centre for Urban Studies

University of Amsterdam
Plantage Muidergracht 14
1018 TV Amsterdam
The Netherlands

Phone: +31 20 525 4081
Fax: +31 20 525 4051

Website: urbanstudies.uva.nl
Email: urbanstudies@uva.nl

The **Centre for Urban Studies** (CUS) houses the Urban Studies Research Priority Area, a strategic initiative of the University of Amsterdam. It brings together urban scholars in sociology, geography, planning, political science, economics, development studies and other disciplines. The Centre supports existing urban research programs and stimulates interdisciplinary collaborative projects. With 39 academic staff and over 60 PhD students, it is among the largest programmes of its kind in the world. The Centre works closely with both academic and non-academic partners and has developed a variety of institutional relations with other leading institutions. CUS is part of the AISSR, the Amsterdam Institute of Social Sciences Research, in the Faculty of Social and Behavioral Sciences at the University of Amsterdam

DO PLANNING SUPPORT SYSTEMS IMPROVE PLANNING OR NOT

A measurement framework and randomized experiment

Marco te Brömmelstroet

Abstract

Although Planning Support Systems (PSS) were, and are still, seen as very promising for improving urban planning practices, several decades of developing generations of such technologies have not bridged the implementation gap. To close this gap, a realistic evaluation of what works and what doesn't is needed. In this paper I present a framework to support such evaluation based on the 'quality of planning' concept. This framework is subsequently used to assess the added value of a state-of-the-art PSS. In a controlled experiment I test if there are systematic differences on planning quality between an assisted planning process supported by a PSS and an unassisted planning process, without a PSS. The results are presented in both descriptive and parametric statistics to provide a rich account of both systematic and non-systematic effects. The outcomes show that the PSS brought a systematic added value to a number of dimensions of the quality of the planning process. It was especially strong in supporting the group process and in providing insights in mechanisms of urban problems to the participants. However, the quality of the outcomes was not systematically improved. These findings raise some interesting questions on what PSS aspire and what they currently achieve. Especially the importance of mediating between the instrument and participants seems an important element. The measurement framework was useful in supporting this analysis and could be further developed an used in research on PSS usability.

1 Introduction

In planning research, it is considered a truism that 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; by whom; at which part of the process (Khakee et al. 2000); and how they should be related to planning actions (Forester 1989; Healey 2007; Nonaka and Konno 1998). It is widely accepted that knowledge does bring added value.

Researchers and practitioners in the field of Planning Support Systems (PSS) started from this notion and have developed a wide range of dedicated technologies that aim to support linking, integrating and creating planning-relevant knowledge in specific parts of the planning process (Brail and Klosterman 2001; Brail 2008; Geertman and Stillwell 2003; Geertman and Stillwell 2009; Geertman et al. 2013). To support these planning tasks, PSS always have both a content and process element. These instruments hold a huge potential for the 'hopelessly complex human endeavour' of urban planning (Couclelis 2005). However, as has been consistently demonstrated in a growing number of studies, their use in actual planning practices is very limited. In his seminal study on this 'implementation gap', Vonk summed up the main shortcomings of current PSS: '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). One of the main conclusions of his research into the underlying reasons behind these shortcomings is that there is a fundamental and consistent gap between the world of PSS developers and the reality of PSS users (see already Lee 1973; Lee 1994; and recently confirmed in Te Brömmelstroet 2010; Vonk 2006). In the light of this apparent mismatch, it is striking that very few concerted research efforts have focused on the need to overcome the implementation gap. A recent meta-study showed that the question of usability of PSS is almost non-existent in research circles, and when present, it is only fragmented at best (Te Brömmelstroet 2013)

It is time for a coordinated effort to move forward on understanding and improving PSS usability. Such an effort should aim to uncover what currently works or does not work in which practical contexts and the reason why, as dictated by the approach of 'realistic evaluation' (Pawson and Tilley 1997). The goal of this paper is to develop and use an analytical framework that supports such practice-oriented evaluation. First, the concept of planning quality is defined into a multi-dimensional framework (section 2), necessary for assessing and comparing the usability of PSS in different contexts. This framework is operationalized in section 3, and used in a controlled experiment to test the general added value of a PSS in a typical planning process (section 4). Structured by the framework, the results of this experiment are interpreted and discussed through both descriptive and parametric statistics. The paper is a first exploration of an experimental research approach to the added value of PSS and it builds on arguments why this has an added value. But due to its explorative nature it closes with an extensive reflection and discussion on the lessons learned on research choices and possible ways forward.

2 Performance of PSS

In a recent meta-analysis of PSS literature, a conceptual framework was introduced that formulated 'improving planning quality' as the core goal of PSS (Te Brömmelstroet 2013).

Based on the literature on knowledge management, process management and group model building, the framework was then translated into thirteen dimensions (Dean et al. 2006; Rouwette et al. 2002). Planning quality is first divided into quality of outcomes and quality of planning processes. There is no clear academic consensus about what distinguishes *good* from *bad* planning or a single concept shared concept of planning quality. Instead there are many competing schools of planning paradigms that all weigh in on this debate (overviews in: Allmendinger 2002; Allmendinger 2002; Campbell 1996; Friedmann 1998). One of the fundamental differences between these schools is a focus on the quality of the planning outcome versus the quality of the planning process, especially in recent contributions that argue for a more participatory and inclusive planning style (see e.g. Burby 2003; Healey 2007). Instead of choosing one side in this debate, here I aim to be as inclusive as possible, but with a focus on the more strategic phases of planning. To do so, Fischler (2012, p. 110) provides a very useful starting point, stating that good planning: convey(s) a clear sense of purpose (a vision, a set of values) and a clear definition of the issue(s) at hand, express regard for diverse community needs, present a realistic action plan focused on a limited number of projects or actions, and contain provisions for monitoring. Good plans marry idealism and realism: they help shape the collective environment according to certain goals for tomorrow and in accordance to the means available today. Good plans also combine sound analysis of problems and artful design of solutions. Without a clear definition of issues and goals, without political support from the powers that be (which, in a democracy, includes the voters and their interest groups), and without a sensible implementation strategy, a plan is not likely to have much impact.

In this light it is useful to distinguish (analytically) between quality of planning outcomes and of planning processes. As argued elsewhere, PSS literature argues that both qualities are improved by using PSS, where each instrument focuses on specific sub elements (Te Brömmelstroet 2013). When hypothesizing about potential added value of their instruments, PSS developers distinguish four main outcome elements: novelty, workability, relevance and specificity. This mirrors insights from the field of ideational output, where these four dimensions are further broken down into sub dimensions. For example, novelty turns into originality and paradigm relatedness, while workability is divided into implementability and acceptability (see table 1).

Much more than coming to the best, most effective or most optimal means to a given end (the focus of Decision Support Systems), the quality of strategic planning phases is much more about individual and shared learning. Fischler (2012) noted that 'planners who do good urban planning in terms of process help to make planning an opportunity for public learning and public deliberation'. In contrast to the relatively well-defined choices in operational phases, strategy making is all about making sense of so-called 'wicked problems' (Rittel and Webber 1984). Most PSS developers argue that their instrument increases the ability of the planning participants to learn and share knowledge with others (Amara et al. 2004; Gudmundsson 2011). The abovementioned metastudy identified a number of distinct elements in these arguments (right column of table 1). These dimensions are also found in recent studies on Group Model Building, which specifically focused on supporting group learning with instruments (Rouwette et al. 2002; Rouwette 2003; Rouwette et al. 2009). Note that the first four dimensions relate to personal learning, whereas the latter five relate to the quality of the group process. Again, some dimensions are further broken down into subdimensions. The letters in table 1 are linked to the statements in table 2 and 3.

Table 1: Multi-dimensional framework for measuring quality of planning and performance of PSS

Planning outcome		Planning process	
Novelty		**Reaction**	
Originality	A	Enthusiasm	J
Paradigm relatedness	B	Satisfaction	K
Workability		Credibility	L
Implementability	C	**Insight**	
Acceptability	D	Insight regarding problem	M
Relevance		Insight regarding assumptions	N
Applicability	E	**Commitment**	O
Effectiveness	F	**Communication**	P
Specificity		**Development of shared language**	Q
Completeness	G	**Consensus**	
Implicational explicitness	H	Consensus on problem	R
Clarity	I	Consensus on goals	S
		Consensus on strategies	T

Cohesion	U
Efficiency gains	V

3 Operationalization

To deploy the conceptual framework of planning quality, I have operationalized it by translating each subdimension into several statements (see table 2), closely following the statements used in the field of ideational output. These statements can then be used to ask external raters (i.e. planning professionals) to assess the quality of planning outcomes.

Table 2: Statements on quality of planning outcome

The planning outcome…	
A1	... is ingenious.
A2	... is imaginative.
A3	... is surprising.
A4	... is novel.
B1	... is radical.
B2	... is transformational.
C1	... can be easily implemented.
D1	... is socially acceptable.
D2	... is legally acceptable.
D3	... is politically acceptable.
E1	... clearly applies to the stated problem.
F1	... will solve the problem.
F2	... is effective.
G1	... can be decomposed into independent subcomponents.
G2	... covers who.
G3	... covers what.
G4	... covers where.

G5	... covers when.
G6	... covers why.
G7	... covers how.
H1	There is a clear relationship between actions and expected outcome.
I1	... is clearly communicated.
I2	... is easy to understand.

The subdimensions of the quality of the planning process were also translated into statements (see table 3). Here, I stayed as close as possible to validated statements used in Group Model Building research. Participants of planning sessions can be asked to respond to these statements, which provides insight in their perception of the experienced planning process.

Table 3: Statements on 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 to the problem.
L2	The 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 the relations between the different elements of 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 the opinions of the other participants about the problem has increased.

N2	I understand how other participants in the session perceive the problem.
N3	The other participants understand how I perceive the problem.
N4	I better understand the proposed solutions of the other participants in the session.
O1	I support most of the results that were drawn during the session.
P1	The process has given me insight into other people's opinions and ideas about the problem.
Q1	During the sessions we developed a shared professional language.
Q2	During the sessions a platform emerged that supported the sharing of ideas.
R1	We have reached a shared vision of the problem.
R2	The results integrated the participant's diverse opinions and ideas.
R3	We were able to reach a consensus on the problem.
S1	We have reached a shared vision on the strategic goals.
T1	We have reached a shared vision on the possible solutions.
U1	I had a strong sense of being part of a group.
U2	The session brought me closer to the other participants.
U3	We experienced conflict during the session.
U4	There was conflict about the task we had in the session.
V1	The session was time efficient.

These two operationalized frameworks are sensitive to a wide variety of potential impacts of a PSS on planning quality. It allows testing of which attributes of planning outcomes and processes benefit from PSS support, and thus supports the realistic evaluation of PSS (Pawson and Tilley 1997).

Research design

The aim of this operationalized framework is to support a wide variety of research settings to assess and improve PSS usability, and to develop insights into the effectiveness of interventions. Here, I will employ the framework to test the general hypothesis that PSS bring an added value to planning processes in a controlled experimental research design.

Why a controlled experiment?

As discussed elsewhere, empirical research on the use of PSS is almost completely based on single case studies performed by the PSS developer him/herself (Te Brömmelstroet 2013). As each PSS is designed and applied in a certain planning context, this approach has provided us with a large body of very rich insights into context specific- and situated applications (overviews:Brail and Klosterman 2001; Brail 2008; Geertman and Stillwell 2003; 2013). However, this dominant research design also puts a severe limitation on developing a general understanding of PSS usability. The context-embedded nature of most empirical research makes it hard, if not impossible, to isolate the effects of the PSS intervention from the effects of the context variables. This limits the internal and external validity of this research and poses serious limits to the desired link up with the existing strong theoretical work (such as: Lee 1973; Batty 1994; Klosterman 2001; Vonk et al. 2005; Vonk 2006; Geertman 2006; Klosterman 2012).

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, p. 5). This systematic effect is associated with (and probably caused by) the manipulation of the causal factor. Ideas about causal factors can be taken directly from PSS theory, and the outcomes can be translated directly back into this theoretical body. An additional benefit of this is that it offers a research vehicle for quick and easy replication to control for the difference in context and its influence.

Use of experiments in PSS literature

A small number of studies in the PSS literature have applied the experimental design. Their main characteristics and shortfalls are shortly discussed below. In a field experiment (Ligtenberg and Vonk 2010) compared two PSS development methods as causal mechanism for the acceptance functionality and usability of a sketch PSS. 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, concluded that the socio-technical development leads to better results. Because treatment and control groups were not independent, the analytic power is limited; the findings are potentially influenced by learning that takes places in between groups.

Nyerges et al. (Nyerges et al. 2006) tested two different set-ups of the same PSS (mainly the process of interaction with the instrument) in two different groups. The process set-up influenced the number of options generated as well as the level of consensus and satisfaction of the planners.

In another study Jankowski and Nyerges (2001) used a laboratory experiment setting. One hundred and nine student volunteers divided in 22 groups were supported by a moderated PSS. Each group received five randomly assigned tasks that involved site selection, but varied on complexity, conflict and access to technology. There was no control group. The findings suggested that the PSS maps did play a role in group decision-making (albeit a very limited one).

A second laboratory set-up is the recent study completed by Arciniegas et al. (2013) with thirty-two students to test four hypotheses on the effectiveness of three different 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 interdependence between the treatments. The authors did acknowledge that learning could have taken place and could have influenced the results (i.e. perceived and observed effectiveness of the three tools). Unfortunately due to the absence of a control group, the study does not give insights into the general added value of PSS for planning.

4 Set-up of the experiment

I designed the study as a randomized, controlled laboratory experiment, with the aim to optimize the internal and external validity of our findings (Bryman 2008). Testing the general claim of the PSS literature (that it improves planning) calls for a strong focus on the ability to connect our findings to the theory. Although I took careful steps to mirror the characteristics of urban planning practice as much as possible, some of the ecological validity of the findings was reduced. Through discussions with the PSS developers, the relative simple setup of the experiment resembles most fundamental characteristics of how this PSS is used in (workshops in) actual planning processes

Student groups

The experiment took place in November 2012, as an obligatory part of a second year course of the Bachelor's Degree in Urban Planning programme at the University of Amsterdam. The 78 student participants were told that they were taking part in an urban planning competition. These students were randomly divided into groups of six. Within each group, each student was randomly assigned one of six planning roles and received information about the plan (see below) that was relevant for their role and the specific goal for the planning session. Each group had students playing out the following six roles:

- Environmental specialist (air quality),
- Environmental specialist (external safety),
- Environmental specialist (sound),
- Urban designer,
- Transport specialist,
- Project economist.

Each of the thirteen groups received the same planning challenge for an infill area in the old harbours 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 aspects of the plan) and were asked to develop a new plan that would solve these issues in a session of 60 minutes.

Figure 1: Brownfield location in the old harbours of Rotterdam

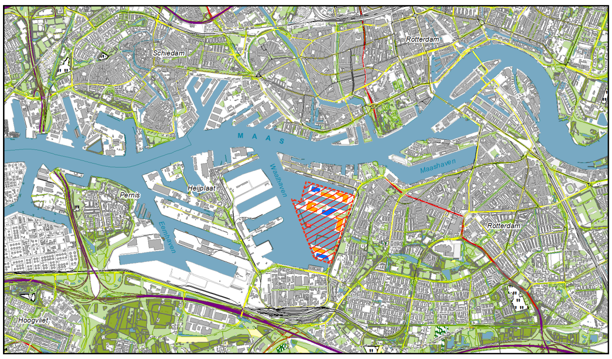
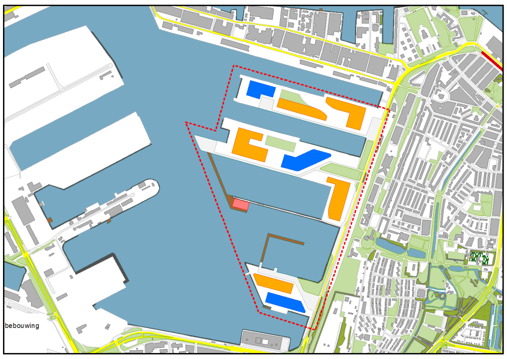


Figure 2: Original design for the area provided to each group



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 maps, instructed to start, and informed when the time for the exercise ended. These six groups worked simultaneously in one room. The treatment groups received the full support of a PSS designed to support these types of planning projects (actually the program was used to support the planning processes for this specific area also in real life). I utilized a state-of-the-art PSS in our experiment, with the following features. Linking to the above discussion on PSS, in this case the treatment makes use of the full support that is offered by the PSS. This means that the treatment groups get both the technology and the support of a mediator and chauffeurs. Without the latter, the first would not be usable in practice.

The PSS links a 2D and 3D design interface with eight different analysis models (transport, noise, air quality, liveability, groundwater, sustainability, external safety and costs modules). Second, it can calculate the effects of interventions for all modules within minutes, so that it can be used in an interactive workshop setting. Third, it uses a surface table to enable participants to interactively engage with the design and effects. Fourth, and particularly important for our experiment, it offers process support by a team of three experts. Two mainly served as guides (chauffeurs) for the application of the PSS. They translated the conversation into interventions in the PSS, ran the models and presented the outputs back to the group. A third person had the role of 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.

Figure 3: Set-up of control group (left) and treatment group (right)

Data gathering and analysis

To find out if there are any systematic differences in the performance of the control and treatment groups I utilized several data gathering techniques.

First, two external planning experts (PhD candidates in Urban Planning at the University of Amsterdam) rated the general quality of the resulting strategies. They were asked to respond statements that referred to the dimensions of table 2. They were not informed of the hypothesis, nor were they aware of which group (control or treatment) had produced which strategy. The raters were asked to respond to each statement by using a 7-step Likert scale, ranging from 'totally disagree' (1) to 'totally agree' (7). This allows analysing the direction and intensity of their attitude towards the planning outcomes (Matell and Jacoby 1971, p. 659).

Second, all participants filled in an evaluation form, sharing their personal perceptions of the process (i.e. their responses to the statements in table 2). And third, I used direct observation by a fourth researcher, by video and audio recording. These observations were used to inform the outcomes of the first three analysis instruments.

For the analysis, the students' responses to the questionnaire were averaged and then compared and tested for systematic differences. To indicate the strength of the differences in effects, I used the p-value of a two-tailed ANOVA f-test to compare two independent means. The statements were grouped and averaged for the sub dimensions and the overall dimensions. The two statements on conflict were first inverted to make them compatible with this process. For the outcome dimensions, the scores of the two evaluators were also averaged and then processed in the same way.

Of the thirteen groups, seven received the PSS support treatment. During the first treatment session, there were extraordinary and severe problems with the process and with the support of the PSS. The PSS mediator was not well prepared and started with a formal presentation. It was immediately clear that this blocked a constructive working process. After this first session, we decided to radically change the process by offering the groups a more open start in which they first developed intervention ideas and then were introduced to the PSS. A quick scan showed that this significantly influenced the results of this group, and thus, for the purpose of this research, the results of this group were excluded from further analysis. Also, five students were absent. This leaves twelve groups and 67 students (31 in control groups and 36 in treatment groups) for the final analysis.

5 Results

This section starts with a descriptive exploration of the performance of all the groups. This gives a first overview of the possible effects of the PSS treatment. After that, I will present the results of parametric test on all the dimensions of table 1, first for the outcome- and then for the process dimensions. After the initial presentation of results, I will discuss the most notable findings.

Descriptive statistics of group performance

To see how the control groups and treatment groups performed, I will present the scores of each group on the main dimensions of process- and outcome quality. All group scores, the average scores and standard deviations are presented in table 4 (control) and table 5 (treatment). Figure 4 and 5 offer a graphic overview of the averages and standard deviations.

Table 4: Process and outcome quality of the six control groups (7 step Likert scale)

Group	1	2	3	4	5	6	**mean**	**stdev**
PROCESS QUALITY	**4,98**	**5,14**	**5,70**	**4,74**	**5,44**	**5,82**	**5,30**	**0,42**
Reaction	4,75	5,10	5,75	4,73	5,65	6,15	**5,35**	**0,58**
Insight	5,08	4,59	5,19	4,25	5,08	5,58	**4,96**	**0,47**
Commitment	5,20	5,40	6,60	5,40	6,00	6,33	**5,82**	**0,57**
Communication	5,20	5,20	5,40	5,00	5,80	5,67	**5,38**	**0,31**
Shared Language	4,50	4,50	5,40	4,50	5,10	4,58	**4,76**	**0,39**
Consensus	5,48	5,64	6,28	5,44	5,56	5,93	**5,72**	**0,32**
Cohesion	5,05	5,50	5,97	4,85	5,50	5,88	**5,46**	**0,44**
Efficiency	4,20	6,20	5,40	5,40	5,00	6,50	**5,45**	**0,83**
OUTCOME QUALITY	**3,60**	**5,00**	**3,27**	**3,83**	**3,40**	**3,67**	**3,80**	**0,62**
NOVELTY	2,25	4,42	2,92	4,17	2,42	3,33	**3,25**	**0,90**
WORKABILITY	5,25	5,88	3,50	3,88	4,50	4,50	**4,58**	**0,87**
RELEVANCE	3,67	5,83	3,17	3,33	3,17	3,67	**3,81**	**1,02**
SPECIFICITY	3,75	4,70	3,55	3,80	3,65	3,55	**3,83**	**0,44**

Table 5: Process and outcome quality of the six treatment groups

Group	1	2	3	4	5	6	**mean**	**stdev**
PROCESS QUALITY	**5,71**	**5,10**	**5,77**	**5,73**	**6,00**	**5,37**	**5,61**	**0,32**
Reaction	5,94	5,15	5,92	6,13	6,10	5,69	**5,82**	**0,37**
Insight	5,38	5,10	5,48	4,98	5,78	5,13	**5,31**	**0,29**
Commitment	6,33	5,67	6,50	6,43	6,60	6,00	**6,25**	**0,35**
Communication	5,33	5,33	5,50	4,86	6,20	5,33	**5,43**	**0,44**
Shared Language	5,42	4,83	4,75	5,50	5,80	4,67	**5,16**	**0,47**
Consensus	6,27	5,00	6,53	6,46	6,48	6,08	**6,14**	**0,58**
Cohesion	5,46	4,60	5,33	5,79	5,45	5,05	**5,28**	**0,41**
Efficiency	5,83	5,00	6,00	6,00	6,20	5,20	**5,71**	**0,49**
OUTCOME QUALITY	**3,46**	**3,83**	**3,15**	**3,27**	**3,08**	**3,33**	**3,35**	**0,27**
Novelty	2,58	3,42	2,17	2,42	2,08	2,33	**2,50**	**0,48**
Workability	4,63	4,25	4,38	4,38	3,75	4,38	**4,29**	**0,29**
Relevance	3,67	3,33	2,67	2,83	3,17	3,50	**3,19**	**0,39**
Specificity	3,45	4,05	3,35	3,45	3,40	3,45	**3,53**	**0,26**

Figure 4: Average scores and standard deviations on quality of the process

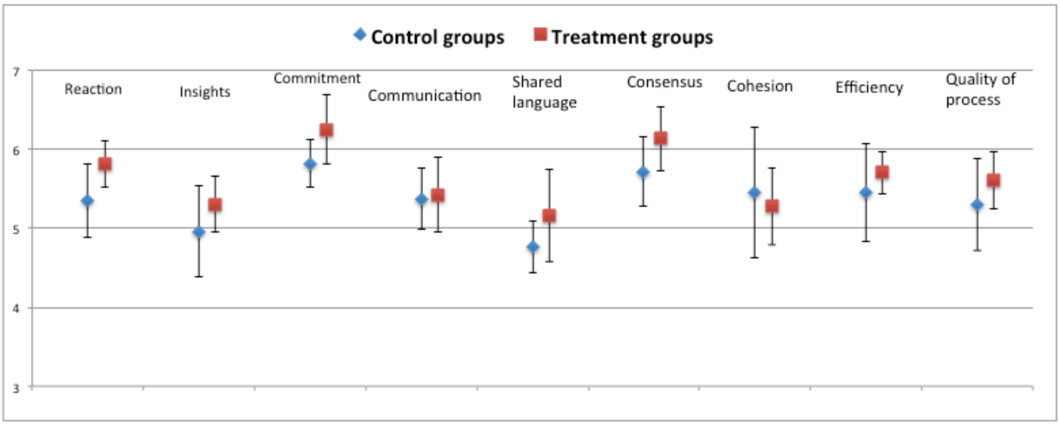
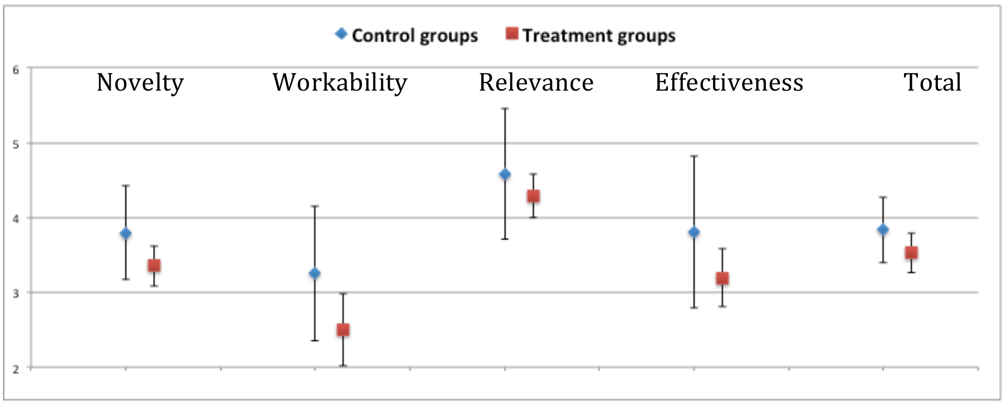


Figure 5. Average scores and standard deviations on quality of the outcomes



Effects on quality of the outcome

To assess if the differences between the control and treatment groups that are presented above are systematic, all dimensions were assessed a standard parametric test (table 6). I will first discuss the outcomes for the quality of the outcome and than go to the quality of the process.

Table 6: Results quality of the outcome

	Treatment averages	Control averages	Differences	P value
N	*6*	*6*		
Quality of planning outcome (total)	3.35	3.80	-0,44	0,143
Novelty	2.50	3.25	-0.75	0.101
Originality	2.42	3.06	-0.64	0.165
Paradigm relatedness	2.67	3.63	**-0.96**	0.061
Workability	4.30	4.59	-0.29	0.455
Implementability	4.25	4.08	0.17	0.813
Acceptability	4.31	4.75	-0.44	0.213
Relevance	3.20	3.81	-0.61	0.199
Applicability	3.33	4.25	-0.92	0.110
Effectiveness	3.13	3.58	-0.46	0.301
Specificity	3.53	3.83	-0.31	0.168
Completeness	3.32	3.66	-0.34	0.123
Implicational explicitness	3.67	4.33	**-0.67**	0.033
Clarity	4.17	4.21	-0.04	0.914

Bold figures are significant on a .10 level

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 f-test for significance. In this ANOVA test the values of all participants are compared. The exploratory nature of this experiment permits using a probability (P value) of .10 to distinguish between systematic and non-systematic differences.

There are only two dimensions that have statistical significant (thus systematic) differences: on implicational explicitness (-0.67) and paradigm relatedness (-0.96). The first effect was contrary to the expectations of the PSS developers. The PSS was expected to support the group in becoming more explicit about the interventions of the strategy. Instead, the opposite occurred. This could have to do with the questions and doubts that were introduced by the instrument. A lower paradigm relatedness is a more ambiguous result: this could indicate that the PSS supported the group in thinking more outside of the box. Breaking with a (ineffective) paradigm can be an explicit aim of PSS, but in this specific case this was not expected. The PSS aims to do the opposite: its underlying models help planning actors to test if their ideas and interventions are within prevailing legal requirements (i.e. following the current paradigm).

The other dimensions also show mainly negative differences, but these are not significant. The low number of groups can explain this result. However, the logic of the measurement framework and experimental setup allows me conclude that I cannot verify an added value (i.e. significantly positive effect) of the PSS on the quality of the planning outcome (see the final sections for a wider debate about how the setup of the experiment influenced this). Although the findings point to an interesting counterhypothesis, they also do not allow falsifying this added value.

Effects on quality of the process

The comparison of the quality of the process is based on the averaged perceptions of the 67 participating students. Again, I use a probability value of 0.1 to distinguish the systematic differences. In contrast to the quality of the outcome, the scores on most dimensions are much higher. This indicates that the participants had a general positive attitude to their experience.

Table 7: Results quality of the process

	Treatment average	Control average	Differences	P value
N	*36*	*31*		
Quality of planning process	**5,61**	**5,32**	**0,29**	**0,084**
Reaction	5.82	5.38	0.44	0.035
Enthusiasm	5.90	5.45	0.45	0.043
Satisfaction	5.97	5.61	0.36	0.068
Credibility	5.61	5.10	0.51	0.033
Insight	5.29	4.98	0.31	0.110
Insight problem	5.27	4.90	0.38	0,102
Insight assumptions	5.29	5.07	0.22	0.298
Commitment	6.25	5.84	0.41	0.093
Communication	5.39	5.39	0.00	0.995
Shared language	5.15	4.76	0.39	0.101
Consensus	6.17	5.73	0.44	0.025
Consensus problem	6.24	5.78	0.45	0.036
Consensus goals	6.06	5.58	0.48	0.030
Consensus strategies	6.09	5.71	0.38	0,102
Cohesion	5.32	5.47	-0.15	0.412
Efficiency gains	5.74	5.48	0.25	0.396

The process quality dimensions also show much more significant effects of the PSS. First of all, there is a systematic effect on the general quality of the process (+0.29). This means that I can verify the claim of a positive added value of the PSS on the quality of the planning process. There are also a number of sub dimensions on which te PSS intervention has a systematic effect: on reaction (+0.44) and all its sub dimensions, on commitment, and on consensus (+0.44) and its sub dimensions consensus on problem (+0.44) and consensus on goals (+0.45). This indicates that the PSS has had systematic positive effects on individual experience and on the group learning process. However, no effects were found on perceived individual learning. Observations indicated that the process moderator and PSS chauffeurs have a strong role to play in these added values. They ensured that the members of each group spent considerable time around the design table; used the opportunity to share individual ideas before going to the map table and collectively interpreted the effects of their ideas.

On the level of the individual statements, the largest negative effect (-0.41) is found in the statement that addressed the sense of being part of a group. Observations confirmed this finding: 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 scored 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. The PSS was very useful in educating the participants on the underlying dynamics behind the sectoral problems that the groups had to address (especially air quality and noise).

6 Conclusions

In this paper I aimed to develop and use an operationalized framework to aid the development of a more generalized understanding of PSS performance. This framework builds on earlier work and research instruments from the field of ideational output and Group Model Building. First, it distinguishes the overall goal of improving planning quality into quality of the outcome and quality of the process. Through this, the framework is sensitive to a wide range of possible effects of PSS and inclusive to the most fundamental ideas on what planning quality is. Both qualities are further divided in subdimensions. To measure the quality of planning outcomes these dimensions were operationalized into 23 statements. External raters were asked to respond to these statements to assess their quality. The process dimensions were operationalized into 31 statements. These were used to collect the perceived quality of the process as experienced by participants.

In the second part of the paper, the framework was applied as a measurement instrument in a controlled experiment. The set-up of the experiment was to test the general premise that PSS (both the technology as the accompanying process support) have an added value for the quality of planning. 73 students were randomly allocated into thirteen groups of six students, and given the assignment to perform an identical planning. By varying the PSS support (PSS-support versus 'business as usual') I was able to analyse systematic effects on planning quality. One group of six students was excluded from the analysis, because of problems with the treatment. Also, five students were absent.

In the experiment I was able to verify an added value for the selected PSS (representing state-of-the-art instruments in the Netherlands) on the quality of the process. I could however not verify a systematic positive effect on the quality of the outcome. Two systematic effects (on paradigm relatedness and on implicational explicitness) were even negative, indicating that the PSS as it was used here might introduce barriers to the quality of the strategies. The lack of verification should not be mistaken for a falsification (i.e. absence of proof is not proof of absence). It could well be that this relates to weaknesses of the research design, research choices or the measurements framework. This will be discussed in the last two sections.

Since the treatment included a complete process and content support (i.e. a PSS in its widest definition), it is hard to point to what exactly caused these systematic effects. I observed that especially the process moderation by the PSS chauffeurs seemed critical. They ensured that the group went to a clear design-analysis-design loop. In the first phase they were instructed to share their personal ideas and design a number of shared interventions. Then, they were invited to the instrument, where effects of these interventions were collectively interpreted. With this, the groups could then sharpen and optimize their original ideas. Even within the short timeframe of the experiment, this seems to increase the perceived quality of the process, especially on personal reaction and group consensus.

One of the most remarkable findings is the large difference between the positive effects on process quality and the negative and absent effects on outcome quality. This is a paradox that deserves further attention in future experiments or case studies, which I will come back to in the discussion section below.

Methodological reflection

The aim of this research was to develop a framework to test claims of the PSS literature and to support a more general learning process towards improving the usability and added value of these instruments. As a first step in this direction, I choose to follow a relative simplistic setup for the experiment. Since I hope that this can offer a direction for wider research initiatives, I want to share some reflections and ways forward.

I am 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. I safeguarded the transparency of the research by providing the details of the operationalization of the definition. As an analytical tool in my experiment, it reflected our observations and was sensitive to a wide variety of potential effects of PSS on planning quality. 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, the internal consistency and measurement validity of the statements need to be tested further.

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

Another question is how closely the experiment followed the golden standard of the experimental research design. Findings on causal relations can only be done 'given appropriate controls' (Goldthorpe 2001). Although I applied randomization in the allocation of students and groups, the relatively small number of participants could lead to some other unforeseen effects due to group composition (e.g., overrepresentation of men or women in groups 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 some of these effects did take place. Role-playing was added to make the group dynamics resemble planning practice better.

Although I measured the added value of a PSS in our experiment, such a research design only indirectly answers the question of added value in planning practices. To find out more about the relevance of our findings, it is necessary to carry out research in real-life planning practices.

Future directions

The observations and experiences yielded many ideas for improvements to the PSS and the process set-up. Also numerous interesting research directions came to the fore during the assessment of the potential added value of PSS in urban planning.

The first direction for further research is to test the interventions that improve PSS usability within the same analytical framework and research design. This would allow the testing of different mechanisms that could improve the usability of PSS, such as improving transparency, flexibility and communication value. Our findings indicate that effective improvement interventions should be sought in improving the knowledge exchange process among individual participants and between the participants and the PSS. One additional focus could distinguish between the effect of the technology of the PSS and that of the mediator.

A second research direction is to focus on the intervening variables that influence how the PSS performs or is perceived to perform. I found that there were significant differences in how different roles perceived the usability of the instrument and the quality of the process. The environmental and transport roles seemed much better supported by the PSS than the urban design role. However, the small number of participants did not allow me to further investigate this. More research is needed to establish the validity of these intriguing suggestions.

Third, the same analytical framework could be used to analyse real world planning settings (with and without PSS) to find out if our findings will be 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 who are closer to real-life planning practice. This could strengthen the external validity. One way to move in this direction is to include students with different specializations within a single experiment.

A fifth direction is refining and improving the elements of the experiment. By having more time per group and simulating more realistic influences (like power, stakes etc), we can improve the link between our findings and our observations in real planning practice.

Sixth, the multidimensional framework could be extended to include more, and more specific, elements of planning quality (i.e. trust between actors). Allowing for this, the generic part of the framework allows cross-case analysis, while the specifics make sure that it also fits very specific PSS aims in very specific contexts.

The last research direction is the possibility to conduct a cross-analysis of the findings. A larger N, obtained after a number of experiments have been conducted, could enable reflection on some general group dynamic and psychological premises (e.g., that overly friendly and comfortable group dynamics limit creative abilities).

Acknowledgement

This research was funded as part of the NWO Sustainable Accessibility of the Randstad research program. Within this program, it was part of the CESAR research project on climate, mobility and urban design issues. I would like to thank all reviewers that have contributed to improving the quality of the research and the reporting.

References

Barlow, J. and Ozaki, R. (2005) Building mass customised housing through innovation in the production system: lessons from Japan, *Environment and Planning A* (37) pp. 9-20

BCJ (2013) Building Centre of Japan, A Quick Look at Housing in Japan http://www.bcj.or.jp/c20_international/agenda/ (accessed 10.10.13)

Beck, U. and Beck-Gernsheim, E. (2002) *Individualization*, *Institutionalized Individualism and its Social and Political Consequences*, London, Sage.

Beer, A. and Faulkner, D. (2011) *Housing Transitions through the Life Course: Aspirations, needs and policy*, Bristol: The Policy Press.

Buchmann M.C. and Kriesi, I (2011) Transition to Adulthood in Europe, *Annual Review of Sociology,* 37, pp. 481-503

Buzar, S., Ogden, P. E., Hall, R., Haase, A., Kabisch S. and Steinfuhrer A. (2007) Splintering urban populations: emergent landscapes of reurbanisation in four European cities, *Urban Studies*, 44(40), pp. 651-677.

Byeon, M-R. (2009) *Single-person households: Changing Seoul*, SDI Policy Report, Seoul Development Institute

Chandler, J., Williams, M. Maconachie, M., Collett, T. and Dodgeon, B. (2004) Living alone: its place in household formation and change, *Sociological Research Online,* 9(3) pp.1-14

Chang, K-S. (2010) *South Korea under Compressed Modernity: Familial Political Economy in Transition*, London and New York: Routledge.

Clapham, C., Mackie, P., Orford, S., Buckley, K & Thomas, I (2012) *Housing Options and Solutions for Young People in 2020*, Joseph Rowntree Foundation, http://www.jrf.org.uk/sites/files/jrf/young-people-housing-options-full_0.pdf

Demographia (2013) 9th International Housing Affordability Survey http://www.demographia.com/dhi.pdf (accessed 10.10.2013)

Dewilde, C. & Raeymaekers, P. (2008), 'The Trade-Off Between Home-Ownership and Pensions: Individual and Institutional Determinants of Old-Age Poverty', Ageing & Society, 28(6), p. 805-830.

Fielding, T. (1992) *Migration and social mobility: South East England as an escalator region.* Regional Studies, 26 (1). pp. 1-15.

Forrest, R. & Hirayama, Y. (2009) The uneven impact of neo-liberalism on housing opportunities. *International Journal of Urban and Regional Research*, 33, pp.998-1013

Fujita, K. (2011). Financial crises, Japan's state regime shift, and Tokyo's urban policy. *Environment and Planning A,* 43 (2) pp. 307–327

Genda, Y. (2001). *Shigoto no Naka no Aimai no Fuan* [The vague uneasiness of work], Tokyo:Chou Korou Shinsha

Hirayama, Y. (2005). Running hot and cold in the urban home ownership market: The experience of Japan's major cities. *Journal of Housing and the Built Environment,* 20 (1) pp. 1–20

Hirayama, Y. (2010) Housing pathway divergence in Japan's insecure economy, *Housing Studies*, 25(6), pp. 777-97

Hiramaya, Y. (2013) Housing and generational fractures in Japan, in R. Forrest and N.M Yip (eds.) *Young People and Housing: Transitions, Trajectories and Generational Fractures*, London and New York: Routledge

Hill, R.C. Kim J.W. (2000) Global cities and developmental states: New York, Tokyo and Seoul, *Urban Studies*, 37 (12) pp. 2167–2195

Hyundai Research Institute (2010) Seven features of single-person household, *Weekly Economic Debates*, vol. 394

Jamieson, L., and Simpson, R. (2013) *Living Alone: Globalization, Identity and Belonging,* Basingstoke, Palgrave Macmillan

Jones, G.W. (2007) Delayed marriage and very low fertility in Pacific Asia, *Population and Development Review* 33, 3: 453-478.

Kim, K.H. (2004) Housing in the Korean Economy, *Journal of Housing Economics*, (13)4, pp. 321-341

Kim, J-H, Choi, M.J. and Ko, J. (2009) Mismatch between homeownership and residence in Korea, *Housing Finance International*, September, pp27-33

Klinenberg, E. (2012) *Going Solo: The Extraordinary Rise And Surprising Appeal Of Living Alone*, London: Duckworth Overlook

KNSO (2013) Korean National Statistics Office, http://kostat.go.kr/portal/korea/index.action (accessed 10.10.13)

KMHLW (2011) Korean Ministry of Health, Labour and Welfare http://english.mw.go.kr/front_eng/ (accessed 10.10.13)

KMLTMA (2008) Korea Ministry of Land, Transport and Maritime Affairs, *A manual of 2008 rental housing*, Gwacheon, South Korea: KMLTMA.

KMLTMA (2010) Korea Ministry of Land, Transport and Maritime Affairs, 2010 Housing yearbook, Gwacheon, South Korea: KMLTMA.

Kubo, T and Yui, Y (2011) Transformation of the housing market in Tokyo since the late 1990s: Housing purchases by single person households, *Asia and African Studies*, XV(3) pp.3-21

Kwon, S, and Holliday, I. (2007) The Korean welfare state: a paradox of expansion in an era of globalisation and economic crisis, *International Journal of Social Welfare*, 16(3) pp. 242-248

Lee, H. (2014) Housing transformations and changing social landscapes in South Korea, in J. Doling and R. Ronald (eds) *Housing East Asia: Socioeconomic and Demographic Challenges*, New York and Basingstoke: Palgrave Macmillan

Leitner, H., Peck, J. & Sheppard, E. (Eds.) (2007) *Contesting neoliberalism: Urban frontiers,* Guilford Press, New York.

Lützeler, R. (2008). Population increase and 'new build gentrification in central Tokyo. *Erdkunde,* 62(4), pp. 287–299.

Machimura, T. (1992) The urban restructuring process in Tokyo in the 1980s: Transforming Tokyo into a world city, *International Journal of Urban and Regional Research*, 16(1), pp. 114–128

McKee, K. (2012). Young people. Homeownership and Future Welfare, *Housing Studies,* 27(6), 853–862

MIAC (2012) Japanese Ministry of Internal Affairs and Communications, Population http://www.stat.go.jp/english/data/index.htm

MLIT (2003) Ministry of Land, Infrastructure and Transport, White paper on the Tokyo Metropolitan Area 2003, Tokyo: National Printing Bureau

MLIT (2007) Ministry of Land, Infrastructure and Transport *Fudosan no Shoken-ka Jitta Chousa* [survey of Real Estate Securitization], Tokyo: National Printing Bureau

MLIT (2010) http://www.mlit.go.jp/hakusyo/mlit/h24/hakusho/h25/html/n1223000.html

MLIT (2012) Ministry of Land, Infrastructure and Transport, *White Paper on Home Ownership*, http://www.mlit.go.jp/hakusyo/mlit/h24/hakusho/h25/pdfindex.html (accessed 08-10-2013)

Mulder, C. H and Billari F.C. (2010) Homeownership regimes and low fertility, *Housing Studies*, 25(4) pp. 527 – 541

Park, B., Hill, R. & Saito, A. (eds.) (2012) *Locating Neoliberalism in East Asia: Neoliberalizing Spaces in Developmental States,* Wiley-Blackwell

Park, Y.J. (1994) The rise of one-person households and their recent characteristics in Korea, *Korea Journal of Population Development*, 23(1), pp.117-29

Ronald, R. (2008). *The Ideology of Home Ownership: Homeowner Societies and the Role of Housing.* Basingstoke: Palgrave Macmillan.

Ronald R. & Alexy A. (2011). Continuity and change in Japanese homes and families. In R. Ronald & A. Alexy (eds.), *Home and Family in Japan: Continuity and Transformation* (pp. 1-24). London: Routledge.

Ronald, R. & Hirayama, Y. (2009) Home alone: the individualization of young, urban Japanese singles, *Environment and Planning A*, 41, pp. 2836–2854.

Saito, A. (2003) Global city formation in a capitalist developmental state: Tokyo and the Waterfront Sub-centre project, *Urban Studies*, 40(2), pp. 283–30

Sato, I. (2007) Welfare regime theories and the Japanese housing system, in Y. Hirayama and R. Ronald (eds) *Housing and Social Transition in Japan*, Routledge, London

Son, I-J, Lee, U-Y., and Lee, I. (2007) A study on the symbolism and fashion of gold miss from the perspective of mass media, *Journal of the Korean Society of Costume,* 57(8), pp. 89-98

Song, J. (2009) *South Koreans in the Debt Crisis: The Creation of a Neoliberal Welfare Society*, Durham, Durham University Press

Song, J. (2010) A room of one's own: the meaning of spatial autonomy for unmarried women in South Korea, *Gender Place and Culture*, 17(2), pp 131-149

Standing, G. (2011) *The Precariat: The New Dangerous Class*, London: Bloomsbury

Tiovonen, T. (2013) *Japan's Emerging Youth Policy: Getting Young Adults Back to Work*, London and New York, Routledge

TMG (2012) Tokyo Metropolitan Government, HP, http://www.toukei.metro.tokyo.jp/tnenkan/tn-eindex.htm (accessed 10.10.2013)

Tsuya, N., Bumpass, L., Choe, M.K. and Rindfuss, R (2005) Is the gender division of labour changing in Japan? *Asian Population Studies* 1(1): 47–68

Waley, P. (2013) Pencilling Tokyo into the map of neoliberal urbanism. *Cities*, 32, pp. 43–50

Yamada, M. (2005) *Meisou suru kazoku: Sengo kazoku moderu no keisei to kaitai [Runaway family: Decline in the post war family model*] Tokyo, Yuhikaku

Yip, N.M. (2013) Homeownership, cohort trajectories and Hong Kong's post-80 generation, in R. Forrest and N.M. Yip (eds) *Young People and Housing Transitions, Trajectories and Generational Fractures,* London and New York, Routledge

Yui, Y (2004) Daitoshi niokeru shingeru josei no manshon konyo to sono haikei [Condominum purchase by single women in large cities] in Y. Yui, h. Kiyama, Y. Wakabayashi, and Nakazawa, T. (eds) *Hataraku josei no toshi kukan* [*Urban Space and Working Women*], Tokyo, Kokin Shouten.