EXPERIMENTING WITH A DESIGN EXPERIMENT

EXPERIMENTANDO CON UN DISEÑO EXPERIMENTAL

Judith Bakker j.h.m.bakker@utwente.nl

Institute for Innovation and Governance studies. The University of Twente. Netherlands.

Bas Denters s.a.h.denters@utwente.nl

Institute for Innovation and Governance studies. The University of Twente. Netherlands.

Abstract

The design experiment is an experimental research method that aims to help design and further develop new (policy) instruments. For the development of a set of guidelines for the facilitation of citizens' initiatives by local governments, we are experimenting with this method. It offers good opportunities for modeling interventions by testing their *instrumental validity* –the usefulness for the intended practical purposes. At the same time design experiments are also useful for evaluating the *empirical validity* of theoretical arguments and the further development of these arguments in the light of empirical evidence (by using e.g. the technique of pattern matching). We describe how we have applied this methodology in two cases and discuss our research approach. We encountered some unexpected difficulties, especially in the cooperation with professionals and citizens. These difficulties complicate the valid attribution of causal effects to the use of the new instrument. However, our preliminary conclusion is that design experiments are useful in our field of study.

Keywords

Causality; Citizens' initiatives; Design experiment; Pattern matching; Policy instrument design.

Resumen

El experimento de diseño es un método de investigación experimental que tiene como objetivo diseñar y desarrollar posteriormente nuevas herramientas (políticas). En este artículo experimentamos con este método para desarrollar un conjunto de directrices que permitan a los gobiernos locales facilitar las iniciativas ciudadanas. El método ofrece la oportunidad de modelar las intervenciones poniendo a prueba su validez instrumental (su utilidad para el fin práctico que se proponen). Al mismo tiempo, los experimentos de diseño son útiles también para evaluar la validez empírica de las discusiones teóricas y el posterior desarrollo de esas discusiones a la luz de la evidencia empírica (usando, por ejemplo, técnicas de concordancia de patrones). En este trabajo describimos cómo hemos aplicado este método a dos casos y discutimos nuestro enfoque de investigación. Hemos hallado diversas dificultades inesperadas, sobre todo en la colaboración de profesionales y ciudadanos. Esas dificultades hacen más complicado que se puedan atribuir de forma válida efectos causales al nuevo instrumento. Sin embargo, nuestra conclusión preliminar es que los experimentos de diseño son útiles en nuestro campo de estudio.

Palabras clave

Causalidad; Concordancia de patrones; Diseño de herramientas políticas; Experimento de diseño; Iniciativas ciudadanas.

INTRODUCTION

Experimentation as a research method in political science and the science of public administration was out of vogue for a long time. As King, Keohane and Verba observe: "Political science is rarely experimental. We do not usually have the opportunity to manipulate the explanatory variables; we just observe them." (1994: 185). In public administration the situation does not seem to be much different (Houston and Delevan 1994; Perry 1994).

Morton and Williams claim that: 'Political scientists often rule out experimentation as a useful method for many interesting research questions' […] However, experimentation is increasing dramatically in political science' (Morton and Williams 2008). Indeed, the number of scientific articles based on experimental research in our discipline's journals has grown over the recent decades. In such articles experiments are often used to test scientific hypotheses in laboratory settings where both the intervention and the contexts are stringently controlled and manipulated. On the other hand, in the world of politics and public administration 'experiments' and 'pilot projects' abound: new institutions (such as the direct election of mayors, new electoral rules), and new modes of citizen participation (interactive governance and more recently citizens' initiatives), are put to a practical test outside strictly manipulated laboratory settings before implementation on a wider scale. Sometimes, but not always, such 'experiments' are followed by evaluative research. But even if such laudable efforts at evaluation are made, more often than not they fail to provide a sound evidence base to determine the instrumental validity of the instrument —to determine if its works.

In our research project 'Citizens making their neighborhood', we use 'design experiment' methodology, following on the ideas of Gerry Stoker and Peter John (Stoker and John 2009). They argue that design experiments are a useful tool in the context of applied research in political science and public administration. John et al. (2011) demonstrate the usefulness of experimental research designs (including design experiments) in developing successful policy interventions aimed at changing civic behavior. In our research project we follow in their footsteps by using a design experiment to develop successful governmental strategies to facilitate so called citizens' initiatives (from here on referred to as CI's). CI's can be considered as a form of citizen participation in which *citizens collectively engage in activities aimed at providing local "public goods or services", (for example the promotion of livability and safety in their street, neighborhood or town).* A distinguishing feature of such initiatives is that citizens decide both on the aims of their activities and play the main role in the actual implementation. The role of the municipality (and other relevant local agencies) is mostly limited to providing support and facilitation. Many Dutch municipalities —partly in the light of the need for budget cutbacks— want to stimulate such CI's, but are looking for guidance and advice on how they can play a facilitative role in such a way as to maximize the chances for success of these initiatives. The success of such initiatives may be defined in terms of a variety of criteria, such as:

- The attainment of citizens' objectives
- Improved quality of the neighborhood
- Empowerment of citizens
- Increased trust in government and politics

Stimulating and facilitating CI's asks for new patterns of interaction between citizens and governments, and requires new know-how on the side of local officials and their organizations. This is now widely recognized, which is testified by a host of reports, brochures and case-descriptions with lists of do's and don'ts in case of CI's, sometimes even advertised as 'best practices'. But a theory-based, validated tool to establish what interventions contribute to successful CI's is as yet unavailable. Therefore, the project 'Citizens making their neighborhood'[1] was initiated. The aim of this project is to develop a set of guidelines that public sector professionals can use in facilitating and supporting CI's. These users include local actors on two levels of action[2]:

- **Frontline workers in** municipalities and other local organizations (e.g. housing corporations or welfare organizations) who support and facilitate participation of residents in the context of urban renewal and neighborhood policies.
- **Managers and (elected or appointed) officials** who bear strategic responsibilities and are also responsible for securing the openness and responsiveness of their organization in case of new modes of citizen self- governance, including CI's.

In this contribution we focus on the development of guidelines for the frontline workers. We worked together with these frontline workers in our research project. In our interactions with them we were aware of the fact that over many years they were expected to look at residents as either *subjects* (to abide by governmental rules and regulations), *clients* (when consuming public services) or *citizens* (voicing their needs and demands through elections and other non-electoral channels for public participation). However, many professionals are not accustomed nor professionally equipped to deal with *self-governing community members*, who initiate improvements in the quality of their neighborhood or other parts of their lives. The guidelines to be developed therefore should help these officials in mapping the motivations, possibilities, problems and frustrations of residents involved in initiatives. This requires that the frontline-workers develop an

---

[1] The project is based on a collaboration among three Dutch cities, the Dutch Ministry of the Interior and Kingdom Relations, NICIS (Netherlands Institute for City Innovation Studies) and two Dutch universities (Twente and Amsterdam).

[2] Rank-and-file officers work at a third level, in the regular departments of the municipal or service organization. They provide standard predefined services, facilities or implement policies and enforce rules and regulations.

adequate understanding of the dynamics of voluntary civic action. That is precisely the reason why our main theoretical frame of orientation is found in the Civic Voluntarism Model formulated by Verba and his associates (1995) and the CLEAR model (Lowndes et al. 2005) that is based on this model.

The development of guidelines for the professionals in the context of a design experiment implies a two stage process. First, in the early stages, an instrument has to be designed. Typically this is done by the researchers based on their theoretical knowledge. But the initial design might also be (partly) based on the professional knowledge and experience of practitioners or the everyday knowledge of citizens. Second, the researchers develop theoretical expectations (hypotheses) about the anticipated effects of the actual use of this set of guidelines (the experiment) and confront these expectations with data collected about the actual process and its outcomes. This second stage is in fact identical to performing a formative evaluation in which the effectiveness of a policy intervention is investigated. Subsequently, if the results of this initial evaluation point to opportunities for improving results, the initial instrument is to be adapted and the improved instrument is re-tested.

In this contribution we first will elaborate on the reasons why we have selected the design experiment as a research method. We do so by comparing it to other methods. We will subsequently describe how we conducted our design experiment. For the various stages of the design experiment we have described our choices and their implications. Finally we reflect on the implications of our choices in the light of a number of (partly unexpected) problems.

CRITERIA FOR SELECTION OF A PROPER RESEARCH DESIGN

We considered a number of criteria in selecting a proper research design. First (1), the research design would have to be theory-based. Our set of guidelines should be based upon well-founded hypotheses about the expected effects of particular interventions. Thus, theory would provide a sound basis for the initial design of the instrument. Theory, moreover, provides the main benchmarks (expected results) against which the observations of actual results are to be measured. If observed results fall short of these theoretical expectations, the theory (or an adapted version thereof) has to provide the basis for recommendations for further development of the instrument. Second (2), the research design should allow to make valid causal inferences (internal validity). The question is: does the instrument (here the guidelines) under study indeed contribute to observed results in the context where the intervention was made? Third (3), the design has to allow for opportunities to intervene (adopt our guidelines for professional behavior) and to control key parameters of the research situation. CI's are typically characterized by continuous interactions between a variety of different actors over longer periods of time. Preferably, the intervention and its effects should be studied in settings that reflect such complex conditions. Fourth (4), the research should preferably allow to generalize con-

clusions to a wider population of cases (external validity). Ideally, the adopted design should score high on all four criteria.

In the next section we will discuss classical experiments, action research and design experiments in the light of these criteria.[3] This comparison will bring out the strengths and weaknesses of the various alternatives.

Classical experiments

In classical experiments an intervention (traditionally called 'treatment') is performed on one group of subjects, whilst a control group does not undergo the treatment. By random assignment of the subjects to the experimental and the control group it is decided who receives the treatment and who does not (randomized controlled trial). A pre-test is administered in both the experimental and the control group in which the situation with regard to the dependent variable is measured. A similar test is also used to measure the situation after the treatment. If a difference occurs in the changes between the pre- and post-tests occurs between the two groups, this difference can be attributed to the treatment.

Obviously, experiments require a firm theoretical basis to identify both the treatment variable and the dependent variable and to formulate the anticipated effects of the former on the latter (1). Therefore the experiment meets our first criterion.

The power of experiments to make causal inferences (criterion 2) is so strong, that this method is called 'golden standard' in research: 'These trials produce a more warrantable form of knowledge than other methods' can provide and for that reason experiments are considered as a superior tool for establishing causal relationships (Babbie 2004; John, Cotterill et al. 2011). Basic conditions for establishing causality are met: the cause and the effect co-vary; the cause precedes the effect and the effect cannot be spurious (because of the randomized control).

With regard to our third criterion (3) the classical experiment design takes place in a laboratory setting which allows for a considerable degree of control over the conditions in which the intervention is made and for control over the intervention itself. Finally, the generalization of results is straightforward, because the subjects in the treatment and the control group are typically individuals, randomly selected from a well-defined population (4).

Nevertheless, classical experiments require rather specific and exacting conditions. First of all, the number of subjects in the two groups has to be large enough to even out all variance in relevant characteristics. As Ray Pawson and Nick Tilly (Pawson and Tilley 1998) argue, however, if subjects are not individuals but collectivities it may be difficult (and practically impossible because of the costs) to find an appropriate number of cases; the possible variance might be very high because of the complexity of the systems,

---

[3] We have only included research designs that allow for some degree of control over the treatment variable (criterion 3). For this reason the most similar systems design (Lijphart 1971; Meckstroth 1975) and the most different systems design (Przworski & Teune 1970; Denters & Mossberger 2006), will not discussed.

and the population to select from small. Furthermore, although an experiment allows for strict controls on the intervention and conditions, this typically is achieved by a strictly controlled laboratory setting. In the context of CI's it will be difficult if not impossible to realize either of these exacting conditions: the number of CI's that can be selected for experimental purposes is likely to be limited and imposing experimental controls in a laboratory setting is impossible.

Action research

The goal of action research is to acquire knowledge for the good of disadvantaged groups in society. "The disadvantaged subjects define their problems, define the remedies desired, and take the lead in designing the research that will help them realize their aims […]. Implicit in this approach is the belief that research functions not only as a means of knowledge production but also as a tool for the education and the development of consciousness as well as mobilization for action." (Babbie 2012:313).

This approach is unique for action research. Just like classical experiments, action research actively and consciously interferes in reality. Here the deliberate interventionism is different from other forms of social science research in which researchers scrupulously try to avoid to influence the process under investigation. Unlike most social scientists action researchers are "… focused on the input of policy makers and the refinements of an intervention. They offer an iterative methodology that relies on cycles of direct interaction with the decision makers and on refining how policy makers intervene" (Stoker and John 2009:362).

Because of their desire for relevance and political preoccupations action researchers are very much practically oriented. Some action researchers even consider theoretical relevance as an obstacle for practical relevance: building on established theories might prevent researchers from asking 'questions that really matter' (Swanborn 1984) in the eyes of their target groups. In the light of our first criterion we conclude that action research is a-theoretical if not plainly anti-theoretical.

In methodological terms it is hard to characterize action research. Neither Babbie, nor Swanborn or John et al. provide a discussion of how researchers operating in this tradition secure the internal validity of their results (criterion 2). Action research has a questionable methodological reputation, because many of its advocates are fervent adherents of the dictum: 'relevance rather than rigor'. Too often the prescriptions of action research are not based on solid empirical evidence. Even Susman and Evered (1978), in general sympathetic towards action research, admit that 'When action research is tested against the criteria of positivist science, action research is found not to meet its critical tests'. However, they claim that another type of knowledge can be generated, and is relevant in terms of generating good organizational knowledge. Action research focuses on specific, real-life, real-time problems and situations, and has an interventionist orientation, aimed at innovation and problem-solving (Susman and Evered 1978). Therefore action research at least to some extent meets our third criterion. Finally, the orientation of action

research on the specific case and its ideologically driven identification with a *particular* target group not only stand in the way of a careful consideration of theory and method, but also might prevent action researchers to consider the generalization of their findings (criterion 4).

All in all, we conclude that traditional action research has been dominated so heavily with political preoccupations that methodological problems regarding questions of internal and external validity were of little concern.

Design experiment

Design experiment methodology combines the action researchers' concern with relevance with an ambition for theoretical and methodological rigor. Historically, design experiments have been the province of the artificial and design sciences, artificial intelligence, architecture, engineering and medicine (Gorard, Roberts et al. 2004:578). Notice that the word *design* in the name of this methodology does not refer to the *research design*. The term design refers to the blueprint of a new *instrument* that is to be developed during the research process. In the social sciences the method was initially used in educational science to help determine what learning strategies and teaching methods might be effective in education. The methodology 'allows the educational researcher to study learning in context, while systematically designing and producing usable and effective classroom artifacts and interventions' (Gorard, Roberts et al. 2004:579). As Stoker & John (Stoker and John 2009) observe, this methodology so far has hardly been used in political science and public administration, while other forms of experimental research are steadily increasing in popularity. The sparse use of design methodology is remarkable. As we have stated in the introduction, 'experiments' and 'pilots' abound in the domain of politics and public administration: there are numerous innovations in public participation, policy programs or institutional reforms that could be studied with the help of design methodology. The design experiment claims to provide an evidence about 'what works' in the early stages of the development of a policy intervention…" (Stoker and John 2009). The aim of design experiments is to provide answers to design issues. To begin with, this implies that in answering its core questions design methodology takes its starting point in available theories. Therefore Stoker and John are right in claiming that [d]*esign experiments are not a-theoretical.* [...] Design experiments work with one or several theories. They develop a practical understanding of what could work , which can derive from social science theory and/or the intentions of those carrying out the intervention. (Stoker and John 2009:360).The theory may serve to provide a priori ideas about certain interventions and their effects. Besides, the experience and professionalism of the officials and researchers involved can suggest new applications of the theory, or can provide new ideas for mechanisms at work.

While the instrument is implemented, the process and the intermediary results are closely monitored and evaluated. On that basis the researcher may give advice on how to improve the instrument. This iterative process of design, trial, error, redesign, trial, etc.

is not only practically relevant. It also provides the opportunity to evaluate the empirical validity of interesting theoretical hypotheses in the context of a field experiment. Our first criterion —a theoretically based method— is clearly met.

One of the most challenging tasks in performing design experiments is to provide a rigorous analysis that allows for maximum feasible internal validity (criterion 2). This is daunting because typically field experiments are conducted in the context of one or only a few cases. Design experiments share with classical experiments the desire to study the effects (based on a before-after design) of theoretically interesting interventions. But design experiments, for reasons explained in the aforementioned, can neither rely on the comparison of an experimental and a control group (composed of randomly selected units of observation), nor is it always possible to use multivariate statistics for demonstrating causal effects and to control for spuriousness.

Instead, design experiments rely on a different strategy for establishing causality. In classical experiments the causal process is considered as a black box and the operation of causal mechanisms is assumed rather than observed. Like other qualitative methods of research, design experiments are based on the identification of causal mechanisms and hypothesize a causal process that is specified in terms of sequences of expected events. On the basis of this, the researcher(s) will collect data on the occurrence of the predicted events, these data units can also be called causal-process observations or CPOs (Collier 2011). This amounts to a careful reconstruction of actually observed sequences of events in the context of one or more experimental situations that are matched with theoretically expected patterns in sequences of events. Therefore this technique is also referred to as pattern matching. Pattern matching is part of the method of process tracing, 'an analytic tool for drawing descriptive and causal inferences from diagnostic pieces of evidence' (Collier 2011).

In addition to the theoretical plausibility of the anticipated effect(s) there are two additional pieces of empirical evidence in design experiments that provide support for the hypothesis that X has contributed to Y (here e.g. the success of a CI). First there should be a change in the scores on Y after the treatment. Moreover, there should be evidence that theoretically plausible causal mechanisms have set in motion sequences of events that have contributed to the observed change in Y. Thus DE-methodology offers the possibility to observe the effects of the use of the instrument over time and of the process taking place after the use of the instrument that may lead to the expected change in Y.

A final concern is the need for external validation of design experiments (criterion 4). In this respect, there are two validity threats. First, design experiments may not be externally valid because of the disturbances introduced in the process by the presence and interventions of the researchers (so-called Hawthorne effects). Second, the external validity of one or a limited number of design experiments is of course a priori limited. Therefore, protagonists of design experiment methodology (Stoker and John 2009) emphasize the need for external validation of the initially developed instrument in a wider number of environments.

Oᴜʀ ᴅᴇꜱɪɢɴ ᴇxᴘᴇʀɪᴍᴇɴᴛ ꜱᴛᴇᴘ ʙʏ ꜱᴛᴇᴘ

On the basis of the above arguments we conclude that for the aims of our research the design experiment promised to be an effective research instrument. But making such a decision in the abstract is one thing. Implementing such a relatively new methodology in practice is another. Therefore, in the second part of this article we will describe our experiences so far in conducting our design experiment.

According to the recent literature (Gorard, Roberts et al. 2004; Stoker and Greasley 2005; Stoker and John 2009; John, Cotterill et al. 2011), a design experiment implies the following steps in the research process. Theory always comes first. Relevant theoretical ideas form the basis for the development of the possible intervention(s) that should lead to a desired outcome. Practitioner's experiences can be used as well, for example, to translate initial theoretical idea's into workable interventions. These interventions have then to be implemented in a small number of relevant real-life situations. Third, intensive and detailed recordings of the project and the intervention(s) are indispensable to allow conclusions about the causal process that links the intervention to the potential effects. If the process or the (intermediate) effects are not in line with the expectations there may be grounds for adaptations, through rounds of iterations; cycles of designs and redesigns. Finally, in order to enhance the external validity the researchers can also perform a study with appropriate statistical power to confirm the effect of the use of the new instrument. In the remainder of this paper we will describe how we went through these different steps in conducting our (still ongoing) design experiment.

Theory first

The aim of this research was to develop an evidence-based set of guidelines for facilitating *successful* CI's. However, one defines the success of such initiatives (see the introduction), an necessary ingredient for success is (continued) involvement of citizens, who initiate and develop these projects. Therefore, we considered stimulating citizens' (continued) involvement as the key factor in defining the success of facilitation of CI's. In order to develop an initial set of guidelines we started our research project by studying available theoretical and empirical knowledge. The civic voluntarism model (Verba, Schlozman et al. 1995) and the related CLEAR-model developed by Lowndes et al. (Lowndes, Pratchett et al. 2006)) provided us with a theoretical basis for identifying the main factors explaining the likelihood of citizens' involvement in CI's. In short, both Verba and Lowndes state that first one must be *able* to participate in order to do so. Time, money and civic skills are needed for people to participate. Second, without motivation no one participates. Only if an individual *wants* to participate, he or she will do so. And furthermore, help from people and organizations contributes to the chances that people participate; networks *enable* participation. Also mobilisation plays a, sometimes underestimated, role; when people are *asked* to participate, the odds that they do raise considerably. Finally, *response* from governmental side is very stimulating. Lowndes abbreviated

these five factors as follows: Can do, Like to, Enabled to, Asked to and Responded to. Together they form the acronym 'CLEAR'.

Knowing the main factors explaining citizen involvement, also provides the basis for considering how professional 'facilitators' might undertake remedial action when problems occur (cf. Lowndes et al. 2006).

On the basis of this theoretical analysis and a subsequent secondary analysis of 21 case studies about CI's (Bakker et al. 2012) we drafted a seven page memorandum with a set of guidelines for professionals involved in the CI's. The memorandum can be considered as a first version of an 'intervention theory'. One of the main principles underlying this memorandum was the presumption that 'facilitators' should not take policies, planning procedures and routines as their starting point, but instead begin at the motivations, ambitions and capabilities of the citizens. This way the CLEAR framework (Lowndes, Pratchett et al. 2006) was made into a useful tool for analyzing the starting situation and finding interventions to support CIs. Our theoretical framework and the memorandum basically served two purposes during the research. On the one hand they provided the concepts and categories that helped us in systematically monitoring the CI's, and the effects of the interventions by the professionals. On the other hand, it enabled us to provide structured feedback to the facilitators and formed the basis for suggestions for adjusting their intervention strategies when necessary. As such this document also served as the first version of the set of guidelines for facilitators of CI's, that was the envisioned 'end product' of our design research. In an appendix we present a checklist which we used to present our analyses and our recommendations. This list is a summary of the tool, and serves here as an illustration of how we worked.

The framework and its underlying 'intervention theory' imply a wide range of hypotheses. These hypotheses link attributes of the facilitator's (orientations and actions) to the key dependent variable (citizens' continued involvement in CI's). The general idea underlying these hypotheses is that through their attitudes, orientations and actions 'facilitators' can affect the various CLEAR factors and via these influence the likelihood of citizens' involvement in CI's. In the context of this contribution it is impossible to provide a complete list of the hypotheses implied in the model. Instead we will use two hypotheses for illustrative purposes in subsequent sections. These hypotheses relate to (one aspect —citizens' skills— of) the Can do factor and the Responded to factor from the framework. The hypotheses read as follows:

> Can do: To the extent that facilitators have an accurate picture of the deficits in the skills that are needed in the group of citizens involved in the CI and supplement any deficits therein, citizens' (continued) involvement in CI's is more likely.
> Responded to: To the extent that facilitators respond to citizens' needs and demands for support in realizing a CI, citizens' (continued) involvement in CI's is more likely.

Experimenting with the set of guidelines will demonstrate whether these theoretical ideas can be confirmed.

Selection of cases

To perform a design experiment, there 'has to be a small number of settings where the experiment is carried out or, in many cases, just one setting'. (Stoker and John 2009:366). A number of considerations of both a methodological and a practical nature were important for the selection of our cases. The cases should satisfy the criteria in our definition of CI's. In addition the projects had to be located in one of the three municipalities participating in the research consortium involved in our research (see footnote 1). Moreover, the project should neither be too short-lived (in order to study changes over time) nor should it exceed over the three year period of the research project. Next to these demands, the willingness of citizens and professionals who are actually involved in the CI's to cooperate in our research was of crucial importance. Especially the professionals' cooperation could make or break the research. Since, after the application of these criteria, there was hardly an abundance of cases to choose from we were happy to find two appropriate cases we could work with. The two selected citizens' initiatives that serve as our experimental cases are the realization of a Neighborhood Museum in the city of Enschede and the 'Street Name Project' in Hengelo. In both of these cases we found professionals who were not only willing to cooperate, but also very interested in the research.

Due to practical contingencies we had to begin our fieldwork for both of these projects before the first version of our set of guidelines was fully developed. In both our cases the initiating citizens, on the basis of their own legitimate considerations determined the starting moment. From a researchers' perspective it would have been better if the groups had started a couple of months later, so that we would have been able to test a prototype of our instrument at the outset of the process. However, for obvious reasons we as researchers were not in a position to determine the starting moment.

Collaboration with policy-makers, practitioners and participants

In literature on design experiments in education usually the role of teacher and researcher are combined in one person. Stoker and John (2009) think that in public administration applications it is very difficult to imagine researchers taking over the role of (for example) a social worker or a police officer. Indeed, in neither of our cases taking over the role of the supporting officials was an option. It is the official, not the researcher, who already has a relationship with the participating citizens, who knows the neighborhood and who is formally responsible for the interventions. A good working relationship and a mutual understanding between the researcher and the professional(s) involved was therefore indispensable. In both cities we worked closely together with the front-line officials concerned with the CI's. We also involved their managers and policy officers to gain support for our recommendations on how the experiment were to be implemented.

The role of the researcher in our project is dual, we made recommendations as to how the front-line professionals were to apply the set of guidelines to each situation and

we observed how the process of realization of the initiatives went on[4]. In the course of the development of the projects, we first explained the intervention theory memorandum and we proposed to the professionals to have two meetings for every round of iterations. One in which we gave feedback of what we observed up to that point. For the second meeting we applied the set of guidelines to the situation on that moment. In doing so, we took the CLEAR-factors one by one and answered three questions:

- What is needed in the case of this CI (with respect to e.g. the 'Can do', and the 'Responded to' factor)?
- What is already available (Which skills, motivations, networks etcetera are available within the group of citizens)?
- And finally: what aspects are problematic (what is still required?) and how and by whom can any of the problems be solved?

On the basis of the responses to these questions we formulated recommendations for possible interventions, and we asked the professionals to implement these.

The cooperation with the participants (the initiators and other active citizens in the projects) was of another nature. When the researchers first met the participants and introduced themselves and the research project, most participants reacted positive, some were indifferent and a few were critical. Considerable efforts were made to explain the position of the researchers and the importance of them attending and observing the meetings. We promised confidentiality and explained our reticence (the observer never commented during meetings on the main process). Notwithstanding our efforts, there was some awkwardness from both sides. Some participants indicated they felt uneasy with the presence of an observer. Others did not bother, or were very curious about the research and the researcher. For the researcher it occasionally was uneasy to only act as an observer, and to refrain from reactions to any of the things that happened and questions that arose. This felt especially awkward at moments when the researcher held information that could have been of good use during the meetings.

Data collection

For the purpose of causal attribution in the design experiment we first needed evidence about the possible effects that CI's might have had on a number of key variables. Here it was especially important to monitor the CLEAR factors (e.g. the resources of and the

---

[4] The researcher in the field has long-term experience in the field of facilitating and supporting active citizens, and is thus well equipped to interpret the interactions between supporting officials and participants, and can make expert judgments on how to translate theory into practical and workable suggestions for interventions.

responsiveness towards the participants), because these factors were critical for the (continued) involvement of citizens in the initiatives (as the key element for CI success). To establish these effects we have first used data about changes in these key variables collected by pretest. In addition to this we wanted to monitor whether any changes in these key variables could be demonstrated to be the result of the CI and the facilitation. For that purpose we used observation reports of meetings—, and additional material, such as policy documents and interviews with managing officials.

Observation reports were written by the researchers based on notes taken during the meetings. In the case of the Neighborhood Museum, there were monthly or sometimes fortnightly meetings with all the volunteers and the community worker. For the Street Name Project—, professionals and volunteers met in various settings, at irregular intervals. We sat in and observed as many meetings as possible. In addition, we tracked the e-mails that were sent between participants and read the documents and minutes (if available) the participants drew up themselves.

For reasons we already mentioned, we were in both cases unable to conduct a genuine pre-test. Instead we conducted interviews with participants at an early stage of each project. Using a standardized questionnaire with additional open questions, we asked all volunteers about their backgrounds, their attitudes towards the municipality and semi-governmental organizations, their supporting officials, as well as their goals and their expectation about support and facilitation. We repeated the interviews with the same individuals near the end of the research period. On the basis of the observation reports, additional materials and the combined results of the pre- and post-tests we were in a position to trace patterns through which interventions by facilitators were (or were not) followed by expected changes in relevant CLEAR factors and contributed to the continued engagement or disengagement of citizens (a key element of CI success).

Applying the set of guidelines. Making interventions

We agreed with the facilitators to apply the set of guidelines in two 'rounds'. In the first round, we used our preliminary instrument primarily to analyze the CI process. On the basis of our observations we made a number of recommendations for each case. In presenting these to the professional(s) we explained how we reached our conclusions, asked the professionals if they agreed on the analysis, and asked them whether they felt it would be feasible to implement the recommendations in the subsequent round (a period of five or six months). Based on these discussions we sometimes adapted our conclusions and advices. But in most instances the professionals agreed with both our diagnoses and the interventions we proposed.

The meetings with the professionals were carefully prepared in our research team in order to minimize the risk that professionals might be offended by our conclusions or recommendations. We wanted to avoid the impression that we were 'breathing down their necks', being judgmental or feeding back information to the professional's super-

visors. Clearly this could have inflicted major damage upon our good working relations with the professionals. For that reason, we phrased all our feedback carefully in terms of observed facts and meticulously explained the theoretical rationale for recommendations, avoiding a critical tone. For example, in the neighborhood museum case we formulated the following advice: "For the progress of this initiative a draft a budget estimate is needed. The participants in this initiative apparently lack the particular skill to make this draft. Something should be done to remedy this. Therefore, we recommend you to intervene. What might help to this avail? Are you willing and able to draft a budget estimate yourself?" Although, sometimes, the professionals might still have interpreted this as a hidden reproach ("Why didn't you help them right away?") our relations with them remained friendly and cooperative.

What we had not anticipated, however, was that in the daily practice of the professionals our recommendations might get "lost". One of the professionals in the Street Name Project during the second round of interviews frankly admitted that, during the six months that had passed after the first talk with him, he had not consulted the documents containing our recommendations. Therefore, only two out of five of the recommended interventions were implemented in this project.

In the second case, the professional involved in the Neighborhood Museum implemented only one of the recommended seven interventions. This recommendation (see our example above) was about the draft of the annual financial year estimate for the group. The other six recommendations were not implemented for various reasons. As a result, we had only limited input for the redesign of our initial set of guidelines. In the next section we will demonstrate how our design experiment provided both confirmations and falsifications of the (intervention) hypotheses based on the CLEAR model.

Evaluating the validity of hypotheses and changing the instrument

In the absence of a strictly experimental design the technique of pattern matching may be an important tool for testing causal hypotheses in the context of a design experiment. In this section we explain how we have used this technique in our study. We will focus on the previously discussed example of the 'Can do' factor (skills) and the related first hypothesis. Based on the hypothesis that a lack of certain skills amongst participants will jeopardize citizens' continued involvement in a CI, we discussed the observed inability of the CI initiators to draft a budget and their lack of knowledge about their financial and legal liabilities, with the facilitator. We suggested that he should intervene and help the group in these respects, by doing it himself. The "intervention hypothesis" underlying this advice was that such support would allow the CI to proceed and boost the motivations and confidence of the participants leading to a continuation of their involvement in the initiative.

This "intervention hypotheses" based on the CLEAR-model imply a particular sequence of events. If the 'Can do' factor is falling short:

> The group will fail to perform crucial tasks (a)→ leading to: a blockage of process (b)→ after having identified the lack of skills as the cause of this (c) → Professional is advised to intervene (d) → The professional heeds this advice (e) the crucial task will be performed (f) → CI proceeds (g)→ Citizens' retain motivation and increase their sense of personal competence (h) → Citizens' remain involved in the CI (i).

The case of the Neighborhood Museum allowed us to observe whether this hypothesis and its implied sequence of events (pattern) matched the actual pattern of occurrences. We observed that in spite of all good intentions and attempts amongst the CI participants, months went by debating financial issues and trying to solve them. Members lacked overview of the revenues they might have, and were not in agreement on spending priorities (a). Moreover, it was unclear who was responsible for budgetary decisions, and what implications this might have for the legal liabilities of the board members (a). Because of this the budget and issues of legal liability and financial responsibilities were a recurrent and major issue on the agenda of the group meetings for an extended period of time, leading to considerable frustration in the group (b). In using the CLEAR framework, we observed that crucial skills were lacking and shared this conclusion with the facilitator (c) and advised him to provide support (d). Subsequently, the professional provided assistance and (after trying to coach the 'treasurer'of the group) made the draft estimate himself. (e). By the professional a draft budget was prepared. It was discussed and agreed upon without much discussion in the next CI meeting (f), this allowed the CI participants to continue their initiative and concentrate on substantive issues, rather than on tedious financial and legal concerns (g) this boosted members motivation to proceed.The new developments that followed increased their sense of personal competence (h) which contributed to their continued involvement in the CI (i).

Our theoretical framework implies many similar expectations. In a previous section we have also provided a second example of such a hypothesis, pertaining to the 'Responded to' factor. Facilitators can have an important role in stimulating the continued involvement of volunteers in CI's by being responsive to their requests and needs for support. Here our guidelines imply that facilitators should be open to requests for support from citizens and take action when such requests are made. If this advice is heeded:

> The group will be supported to perform crucial tasks (a)→ Citizens' therefore will be satisfied or become even more motivated and will increase their sense of personal competence (b) → Citizens remain involved in the CI (c).

In our case studies we have found numerous instances that corroborate this advice. But, surprisingly we also found a number of cases where this was not the case. On the basis of this evidence, we were forced to reconsider our initial guidelines, leading to a reformulation of our instrument. In working with the guidelines, however, we noti-

ced that in at least two situations —notwithstanding the authentic efforts of professionals— there was sometimes considerable frustration amongst the citizens about the responsiveness of their facilitator. In these cases the facilitators acted in line with the idea's in the instrument and did their best to be open to requests for support and provide support when such requests were made. To their surprise, however, the expected benign results did not follow on the requests for assistance. Instead the participants showed quite some resistance. Only on the basis of interviews with participants we began to realize that explicit requests and demands are formulated (or not formulated) *and should be understood*, in the context of citizens' expectations regarding the *role* of the facilitator. On the one hand, this implies that if participants explicitly ask for support they oftentimes not only expect *some kind of response* but may also expect a *particular type of response*. In the case of the Neighborhood Museum, for example, the facilitator upon requests for assistance, tended to take the lead and made decisions himself, to the frustration of the citizens' who were merely expecting his advice on how to do particular things themselves. In other instances, however, expectations with regard to the role of the facilitators may not be stated explicitly. According to our initial guidelines we stated: if facilitators are not asked for support they should refrain from action (even more so, because citizens might be irritated by facilitators who impose themselves; see previous case).

In the Street Name Project, however, citizens' considered it as the self-evident, "natural" role of the facilitator to take a lot of responsibility in the implementation of the plans they had initiated. The professionals had asked them to come up with new initiatives, why should they also execute them? —according to one of the members during our interviews. In the same time the facilitators were vainly waiting for the citizens' explicitly formulated requests for support. On the basis of these observations, in the first round of our two experiments, we added a key element to the set of guidelines: facilitators should not only be open and responsive towards explicitly formulated requests for support but also be aware of the (implicit) expectations that citizens have regarding their role and the style of supporting and types of support provided with the CI participants. On this basis we amended the initial guideline and advised the facilitators to explicitly discuss such expectations with citizens.

These two examples demonstrate how the technique of pattern matching can at the same time be used for evaluating the instrumental validity of policy instruments (here our guidelines for facilitators) and establishing evidence for the empirical validity of theoretical hypotheses (implied in the CLEAR model). In the above we have tentatively outlined some expected patterns (sequences of events) and for illustrative purposes provided some (sketchy) evidence of empirically observed patterns. Using data from our pre-tests, post-tests (e.g. sets of standardized survey items about citizens sense of personal competence), in combination with interviews and other qualitative data can and will be used for a more rigorous and systematic form of pattern matching.

Pitfalls and learning points

'Practice makes perfect', or so they say. We are convinced that the design experiment is as yet by no means a fully developed research methodology. Many of its do's and don'ts are still to be worked out. The experiences of scholars who actually use this new methodology provide an important mechanism through which the methodological quality of design experiments can be improved. It is for this reason that we reflect on some of the lessons we learned in experimenting with a design experiments as a research tool in political science and public administration.

A first important lesson we drew from our project is that the problem of controlling the experimental environment in the context of a politico-administrative design experiment may be even more difficult than in an educational context. In educational settings, one can rely on a relatively stable environment both in terms of timing (rosters), locations (classrooms in a particular building) and groups of students. In CI's, however, the researcher has no control on who initiates a project, why, when, what about, how long and how. All this is in the hands of the initiating citizens. The relevant interaction of the initiators amongst themselves and with other actors can occur pretty much everywhere and at every moment, in changing settings. As we already saw in our project this created problems in organizing timely pretests at the outset. But there were other problems as well. The Street Name Project, for example, started out with three volunteers. One of them did not show up any more after the first encounter. The second volunteer after four meetings announced that he was forced to stop due to health problems. Finally, the third volunteer left unannounced to hibernate in Spain. Although the last person returned and still plays a major role in the process, during a considerable period it was uncertain, whether our carefully selected case was still "alive". These complications came on top of some of the well-known problems of design experiments in education, like having to work with busy professionals, whose priorities may shift over time and who may change jobs without being immediately replaced. Although we only had the opportunity to study two situations closely, we believe that in these respects our cases were not atypical. We suspect that such instability may very well be part of the natural context of CIs.

Secondly, we may have underestimated the complexities created by the multiple prescriptions implied in our instrument. Rather than one singular treatment, our set of guidelines was a multifaceted heuristic device for the professional who would have to use this for developing a coherent facilitation strategy that typically comprised of several interventions. In a context where multiple interventions are likely to be made, the attribution of causal effects (a major challenge in any case) will be further complicated. The design-experiment method might be less demanding when the instrument to be developed would have involved a smaller range of interventions.

In hindsight, we could have improved our research by involving the professionals even more closely, and in an earlier stage. So far, we have made the analyses of the appropriate interventions ourselves, following the initial version of the set of guidelines, and we formulated the recommended interventions ourselves. We now know that at

least one of the professionals was not fully aware of the importance (with respect to the research process) of actually following up on the recommendations; clearly no observations on the effects of the instrument can be made if interventions are not carried out. In addition, we could have made better use of the expertise of the professionals in making adjustments to the tool.

To improve the instrument, newer and better versions should be tested in sequence. Had we had a more advanced version to start with —one that could have been used by the professionals with little or no help and recommendations— a further improved version could have been used in the second rounds of iterations. We would have had a better opportunity to make more advanced improvements before the end of our research project. Up until now we performed two rounds of testing and (partial) improvements; hopefully we succeed in doing yet another round before the end of the project period, that would bring us a step further in the development of an effective tool.

Conclusions

Conducting a design experiment in politico-administrative settings is relatively new and, as a result, proved somewhat adventurous. Applying this method in two cases, we encountered some unexpected difficulties.

On balance, we are convinced that —notwithstanding such problems— the design experiment method has so far helped us to develop a theoretical basis for the design of a practical tool for the facilitation of citizens' initiatives. This list of guidelines is not fully developed yet, but the experiences so far are encouraging. The professionals involved in the project have already indicated in what way the instrument is helpful for them. Moreover, the first effects we have been able to establish provide confirmation for the basic principles underlying the instrument and essentially confirm the instrumental validity of the set of guidelines. When we have collected and fully analyzed new additional data (e.g. the posttests that will still have to be administered) we can make full use of our data and will be well positioned to further improve the instrument.

An important analytical tool used in establishing the impact of particular types of facilitation (X) on the success of citizens' initiatives (Y) was the technique of pattern matching. This technique provides an alternative for traditional experimental and quasi-experimental techniques for causal analysis in the context of small N research. It is based on the identification of key theoretical mechanisms that causally link causes (X) and effects (Y) and the specification of a theoretically expected course of events. This course of events is then compared with the empirically observed course of action. This implied the need to make rather precise expectations about the anticipated effects of the use of the set of guidelines. On the basis of these anticipated effects we were not only in a position to test but, if necessary, also to alter the hypotheses (and redesign the instrument). In due process, this also offered opportunities to explore new insight in the dynamics of the participatory processes in CI's such as the role of expectations with regard to the

supporting officials and how the type of motivations play a role in starting and quitting the participation in a CI.

The external validity of small-N designs (including design experiments) is known to be more problematic than in large-N research. Of course, in our experiment on two cases, this is no different. This design experiment can be considered as a first step; once our instrument is fully developed, the possibility opens up of performing a (quasi-)experimental study with statistical controls.

So far, we have not found a better way to address our specific research objectives. With all its flaws, the design experiment method allows for the development of new instruments, since it allows us to judge its effectiveness and make adjustments. This helps us to also test and adjust theoretical idea's, not in the laboratory, but in a real-life environment.

References

Babbie, E. R. 2012. *The Practice of Social Research*. Independence: Wadsworth Pub Co.

Collier, D. 2011. "Understanding Process Tracing." *PS: Political Science and Politics* 44(4):823-830.

Gorard, S. and K. Roberts, et al. 2004. "What kind of creature is a design experiment?" *British Educational Research Journal* 30(4):577-590.

Houston, D. J. and S. M. Delevan. 1994. "A comparative assessment of public administration journal publications." *Administration & Society* 26(2):252-271.

John, P., S. M. Cotterill, A., et al., Eds. 2011. *Nudge, Nudge, Think, Think: Experimenting with Ways to Change Civic Behaviour*. New York: Bloomsbury academic.

Lowndes, V. and L. Pratchett. 2006. "CLEAR: Understanding Citizen Participation in Local Government and How to Make it Work Better." *Local Governance Research Unit,* De Montfort University Leicester, United Kingdom.

Lowndes, V., L. Pratchett, et al. 2006. "Diagnosing and remedying the failings of official participation schemes: The CLEAR framework." *Social policy and Society* 5(2):281.

Morton, R. B. and K. C. Williams. 2008. *Experimentation in Political Science*. Oxford: Oxford University Press.

Pawson, R. and N. Tilley. 1998. "Caring communities, paradigm polemics, design debates." *Evaluation* 4(1):73.

Perry, J. L. 1994. *Research in Public Administration*. Greenwich: JAI Press, Inc.

Stoker, G. and S. Greasley. 2005. "The case for an experimental approach in applied social research: an illustration from the area of civil renewal policy." IPEG working paper, University of Manchester http://www.ipeg.org.uk/staff/greasley/index.php.

Stoker, G. and P. John. 2009. "Design experiments: Engaging policy makers in the search for evidence about what works." *Political Studies* 57(2):356-373.

Susman, G. I. and R. D. Evered. 1978. "An assessment of the scientific merits of action research." *Administrative Science Quarterly* 23(4):582-603.

Swanborn, P. 1984. "Methoden van Sociaal-Wetenschappelijk Onderzoek [Research Methods in the Social Sciences]." Boom: Meppel.

Verba, S., K. L. Schlozman, et al. 1995. V*oice and Equality: Civic Voluntarism in American Politics*. Cambridge: Cambridge University Press.

JUDITH BAKKER Drs. is junior researcher and Phd-candidate in the Department of Public Administration in the School of Management and Governance of the University of Twente. As a practitioner —consultant, trainer and coordinator— she supported volunteers and active citizens for over eighteen years. She studied sociology at Radboud University in Nijmegen and graduated in 2009. Currently she works for the research project 'Citizens making their neighborhood' and on her Phd-dissertation. Studying the support of active citizenship means for her studying what she used to do.

BAS DENTERS Prof. dr. is prof. of Public Governance and program leader of the research program Innovation of Governance at the Institute for Innovation and Governance Studies at the University of Twente, The Netherlands. His research interest is in urban policy, urban democracy and citizen participation. Until 2010 he was convener of the Standing Group on Local Government and Politics (LOGOPOL) of the European Consortium for Political Research (ECPR). Currently he is member of the Board of the European Urban Research Association, Overseas Editor of Local Government Studies and member of the Editorial Board of Urban Research and Practice.

APPENDIX

Required elements for every Citizens' initiative	What is needed in the project?	What is already available?	Suggested intervention
Resources (CAN DO) Time / money / skills / accomodation	*The ability to draft a financial year estimate*	-	*Supporting professional makes a proposal for a financial year estimate*
Motivations (LIKE TO)			
Networks (ENABLED TO)			
Mobilisation (ASKED TO)			
Response that matches expectations (RESPONDED TO)			

Part of the preliminary instrument: Scheme based on the CLEAR framework (Lowndes and Pratchett 2006).
In Italic: first example as used in the article.