

## RESEARCH ARTICLE

# How evidence becomes authoritative in public policy implementation. Lessons from three Dutch white ravens

J.A.M. Hufen<sup>a</sup> and Joop F.M. Koppenjan<sup>b\*</sup>

<sup>a</sup>QA+ Research and Consultancy in Leiden, Niels Bohrweg 11, 2333 CA Leiden, The Netherlands;

<sup>b</sup>Department of Public Administration, Faculty of Social Sciences, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

(Received 14 November 2012; accepted 14 August 2013)

Policy scientists and public policymakers are continuously struggling with the troublesome relationship between knowledge and policy. On the one hand, policy and policy processes are generally recognized as having a political nature because they prioritize the generation of *will* rather than *knowledge*. On the other hand, it is difficult to accept that knowledge is being wasted or used strategically. One of the main challenges for public policymakers is to reconcile the political rationality of policy processes with the scientific rationality of knowledge and research, for instance by intelligent organization of knowledge generation and knowledge use. This article aims to contribute to the debate on this topic by comparing three policy implementation processes in which knowledge played an authoritative role. On the basis of this analysis of these three ‘white ravens,’ lessons are drawn regarding the conditions under which knowledge becomes authoritative in policy implementation practices.

**Keywords:** knowledge and policy; evidence-based policy; knowledge production; policy implementation

### 1. Introduction

In our current network society, the average education level is high, knowledge and resources are distributed among a large number of organizations, the media offer platforms for critical review of scientific results and citizens have access to expertise through new social media like the Internet. As a result, scientific knowledge, experts and knowledge institutions increasingly have difficulty gaining authoritativeness and influencing policies. An illustrative example is the authority crisis faced by climate scientists when the reports published by the International Panel on Climate Change in 2007 were found to contain inaccuracies (Berkhout 2010). At the same time, politicians and policymakers are advocating the use of knowledge and scientific evidence. The evidence-based policy movement has surfaced in the wake of the New Public Management reforms that have inundated the public administration landscape in the past decades. Politicians and policymakers strive for a more goal-oriented and effective policy practice. Their ambition is a policy that is evidence based (Solesbury 2001; Pawson 2006). Evidence-based policy is based upon objective knowledge – knowledge that has been generated through sound scientific methods (Banks 2009). Evidence refers to knowledge about

---

\*Corresponding author. Email: [koppenjan@fsw.eur.nl](mailto:koppenjan@fsw.eur.nl)

‘what works’ (best practices) that can be deployed in comparable policy situations in a uniform manner (Davies, Nutley, and Smith 2000). This perspective combines a positivist notion of science with an instrumental outlook on use.

The observation that, despite these ambitions, expertise and scientific knowledge have difficulty finding their way into policy is not new. Nevertheless, the problem seems to be becoming increasingly pressing. Theory about knowledge and knowledge-related policy offers a number of explanations for the troublesome transition of scientific knowledge to the actual application in policy formation and implementation processes.

- *Policymakers’ openness towards knowledge.* The earlier debate about knowledge use in policymaking has led to the conclusion that an instrumental view on knowledge use cannot be reconciled with the political nature of policy processes (Weiss 1977; Wildavsky 1979; Patton 1997; Radin 2002). In the arenas in which politicians, public administrators and stakeholders try to achieve their policy preferences, knowledge is a source of power, and research is used to legitimize policy. Knowledge that supports policy intentions is embraced. If knowledge is inconvenient to policymakers’ preferences, however, policymakers will ignore research results, interpret them in their favour or use them selectively. Will and knowledge are sometimes in conflict (Wildavsky 1979; in ‘t Veld 2000; Flyvbjerg 2008).
- *The network-like nature of the policy formation and implementation process.* In our current, complex society, policy is often made and implemented in networks involving various governmental organizations and societal stakeholders. These policy networks of autonomous, yet mutually dependent actors increasingly have a loosely coupled nature, often lacking a shared perception of the problem at hand (Mazmanian and Sabatier 1981; O’Toole 1988). If one of these parties supports its policy preferences with scientific research or invoked expertise, the others are not easily convinced (Gibbons et al. 1994; Jasanoff 1994). Knowledge monopolies traditionally held by, for instance, scientific institutions such as the Dutch Institute for Public Health and Environment are increasingly questioned. Parties then tend to invoke counter-expertise (Koppenjan and Klijn 2004). Consequently, experts and researchers engage in policy advocacy practices that lead to knowledge conflicts and debate. Conflicting truth claims, knowledge conflicts and ‘report wars’ may be the result. This accumulation of ‘contested truths’ makes it difficult for parties to recognize what is true and which knowledge claims and experts are trustworthy (Hoppe 1999; Sabatier and Jenkins-Smith 1993). In the case of ‘wicked’ problems, evidence is always ambiguous, and research does not reduce uncertainty, but adds to it (Noordergraaf 2000; de Bruijn and Leijten 2010).
- *The nature of the process of knowledge generation.* The authoritativeness of knowledge is openly questioned, and not just because high-quality research becomes subject to power games or because opportunistic researchers succumb to ‘policy advocacy’ (Sabatier and Jenkins-Smith 1993; Wildavsky 1995). Another explanation lies in the nature of knowledge generation itself. The positivistic notion that scientific knowledge is the result of objective data and that it produces indisputable truths does not reflect the actual process of knowledge generation (Nowotny, Scott, and Gibbons 2001; Polanyi 2009). All research has its limitations. After all, choices are made with respect to focus, definitions, conceptual models, methods and data interpretation (in ‘t Veld 2000). This

observation has various implications. First of all, the quality of the evidence may be problematic and therefore has a negative impact on its use. Simple scientific models are not necessarily applicable in complex environments (Collins and Evans 2007). In addition, it may be that research or evidence is not conclusive. If institutional monopolies are weakened, results of scientific research may well diverge or experts may disagree on interpretations, methods, or facts. If evidence is not conclusive or if actors have doubt about its quality due to the non-transparency of the knowledge production process, its authoritativeness is at stake.

- *The availability of alternative knowledge sources.* Finally, it is important to note that not all knowledge is scientific by nature, or comes from experts. Different knowledge sources exist. In addition to hard, scientific evidence, there are softer forms of knowledge derived from various sources, including the tacit knowledge of politicians and public administrators, the local knowledge of policy implementers, and the commons knowledge of social networks of citizens, stakeholders and target groups (Hess and Ostrom 2006; Head 2007). Social media also provide parties with knowledge that may compete with that of experts and knowledge institutions. If parties have access to alternative knowledge sources, the truth claims made by experts and knowledge institutions are vulnerable (Bijker, Bal, and Hendriks 2009). This is particularly true if they are based on positivistic notions about science, and if they claim to present absolute truths.

In short, the authority of knowledge and research in processes of policymaking and implementation is no longer self-evident. This, however, does not reduce the relevance and topicality of the ambition to base policy on expert knowledge and the findings of scientific research. The main question is therefore *how* knowledge and research can become authoritative in complex policy environments. This contribution addresses this question by analysing three Dutch cases in which knowledge succeeded in becoming authoritative in the processes of policy formation and implementation. These case studies deviate from the standard idea that policies are the result of political struggle and often lack solid empirical fundamentals. In the Netherlands, the expression ‘white raven’ is used to refer to an exceptional phenomenon. Therefore we consider these cases to be ‘white ravens’.

## 2. A comparative case study of three white ravens

In this article, we focus on the aforementioned question using three case studies in which policymakers in the Netherlands substantiate their policies with extensive scientific research. The knowledge used in these policies has a more than average scientific basis and in that respect these cases can be considered atypical. But next, these policies have to be implemented in complex setting in which various actors are involved. Whether or not the cases will really deserve their status of ‘white ravens’ will depend upon the extent to which the knowledge on which the policies are based will gain authoritativeness in the policy implementation process.

We investigate the factors that influenced the authoritativeness of the knowledge on which the policies studied were based in the arenas of policy implementation, and the lessons that can be learned from these cases about the role of scientific knowledge and research in policy making and implementation.

The following cases are analysed:

- (1) the objective dispensing model that allocates government budgets for social assistance among municipalities as described in the Law on Social Security and Work (2004–2009);
- (2) the regulation on the discharge of warm water as described in the Law on Pollution of Surface Water (2005–2009);
- (3) Central government's policy aimed at energy conservation in houses based on the European Directive Energy Performance Building Sector (2008–2010).

The scientific knowledge used in these cases was expressed in a model, involving in two cases a multi-linear regression model (objective dispensing model and energy performance model) and in one a simulation model (discharge of warm water). Hereafter, we use the terms *model* or *models* when referring to the scientific knowledge used in the selected cases. The outcomes of the policies are to a very large extent determined by the characteristics of the model.

Therefore, the scientific knowledge in the case studies is considered authoritative if the public, semi-public and private actors involved in the policy implementation process accept the characteristics and the outcomes of the model. The characteristics of the model are model assumptions, definitions, framework conditions, the choice of variables, the relations between variables, the operationalization of the variables and the empirical basis of the variables. A model is not authoritative if actors in the implementation process criticize the model, oppose its use, and propose or pursue alternative policies, or if they reject the outcomes of the model, for instance, by starting administrative or legal proceedings.

The empirical basis of the three case studies consists of literature research and interviews. Literature comprises documents on the development and implementation of the policies and the models, the minutes of the parliamentary debates about the policies, media reports and policy evaluations. For the first case study, 16 interviews were held with researchers (who developed the model), municipalities, Association of Dutch Municipalities, the Ministry of Social Affairs and Employment, the Council for Work and Income and the Inspectorate on Work and Income. The second case study was based on 20 interviews with researchers, representatives of industrial enterprises and the energy sector, the authorization organizations and industrial and energy firms. For the third case study, 15 interviews were held with representatives of the knowledge organization Instituut voor Studie en Stimuleren van Onderzoek (Institute for Study and Stimulation of Research, ISSO), the building sector's trade organization (Bouwend Nederland) and the installation sector's trade organization (Uneto-VNI) and the public–private partnership, More with Less, which implements a national covenant about energy saving and sustainable energy.

The fact that the introduction of these models had distributive advantages and disadvantages for the actors involved was not unimportant. The knowledge-based models created winners and losers, and were therefore potentially conflict generating. Because of these political sensitivities, it was to be expected that the policy implementation arenas in which the models were used would be heavily 'evidence resistant' (Head 2007; de Bruijn and ten Heuvelhof 2002).

From a methodological viewpoint, these case studies therefore can be seen as least likely cases and thus critical (Patton 2001; Flyvbjerg 2006; Yin 2009). In the debate on case study methodology, this pattern is known as the 'Frank Sinatra inference' – if you can make it there, you can make it anywhere: if scientific knowledge in these case studies

survives the potential conflicts, it is probable that it will do so in other, less conflictive arenas too (Levy 2008).

This way of selecting cases might suggest that we follow a ‘co-variation analysis approach’, aimed at testing how an independent variable influences outcomes (George and Bennett 2005). However, our main cause is not so much to test a hypothesis, but rather to investigate sequential and situational combinations of causal conditions by pattern matching to arrive at explanations of outcomes in an inductive way. We would qualify our comparative case study design therefore as a qualitative ‘causal process tracing’ approach (Blatter and Haverland 2012).

In Section 3, we present the case studies. For each of the cases, we describe (1) background information on the policy (2) the characteristics of the model developed during the process of policy formation and (3) an overview of the application of the model during the implementation of the policies, establishing the extent to which, and by which processes, the models become authoritative. In Section 4, we discuss similarities and differences between the three cases, looking for factors that can explain whether knowledge became authoritative or not. In Section 5, we present the lessons drawn from this comparative analysis with regard to how knowledge may become authoritative in policy implementation processes.

### 3. Three white ravens presented

#### 3.1. Case study 1: Law on Social Security and Work (WWB)

##### 3.1.1. Background

On 1 January 2004, the Law on Social Security and Work (WWB) was introduced, as a result of which the old declaration system was replaced by the allocation of resources based on the *objective dispensing model*. The objective dispensing model is an econometric model that calculates the number of assistance allowances in a region with parameters such as the percentage of employees in the working population and the percentage of low-income groups aged between 15 and 64 years. The results of the model determine the distribution of budgets for income support among Dutch municipalities. The distribution incentivizes municipalities to guide benefit recipients towards paid labour. If municipalities disburse more than necessary according to the model, they have to make up the rest from their own resources. If they disburse less than their allocation under the model, they have a budget surplus that they can use for other projects to encourage benefit recipients to return to work.

##### 3.1.2. Development of the objective dispensing model during policy formation

The new financial system (objective dispensing model) was developed in collaboration with the Ministry of Social Affairs and Work, implementation organizations of social security laws, municipalities and other stakeholders. The model itself was developed by universities (Erasmus University), semi-public research organizations (Dutch Economic Institute) and private research organizations (APE 2003, 2006; SEOR 2005). Aspects of the model, such as the choice and the operationalization of variables and the choice of databases to be used, were discussed among the stakeholders over the years.

The exchange of ideas concerned not only the question of whether the objective model predicted the number of assistance allowances in a region accurately, but also the consequences of the introduction of the objective model in all Dutch regions. Thus, the

implications of the new model were known before introduction. The budgets of some municipalities were going to increase (advantage municipalities) and those of others were going to decrease (disadvantage municipalities).

The budgetary consequences of the objective model were evaluated *ex ante*, as also the plausibility of the variables, the transparency of the model, the stability of the model and the relevance of the variables. After several versions of the objective model were discussed, one version was chosen. The idea was to introduce the model step by step. After some years, the objective dispensing model would completely determine the allocation of the national budget among local governments. During the development of the model, attention focused on its detailed specifications, and no serious alternative models or instruments were considered.

### *3.1.3. The application of the objective dispensing model during policy implementation*

An evaluation of the implementation of the new law (WWB) reveals that local political leaders and other stakeholders agreed about the vision, goals and operation of the new law (Rapport Meccano 2007). The introduction of the law including the objective dispensing model was widely seen as a success, and the model was accepted as an important building block (Rapport Meccano 2007). The main components of the model were accepted by the ministry, implementing organizations, municipalities and others. There was no need to consider policy changes, although on the initiative of the Ministry of Social Affairs and Employment some aspects of the model were discussed and adapted (definitions and data collection). In a few years, the objective dispensing model had become the undisputed basis to allocate the social assistance budget among municipalities.

Municipalities not only accepted the model as the mechanism for allocation but also tried to gain some benefits by reintroducing welfare recipients to the labour market. National government developed a toolkit of measures that local government could use to reduce the number of allowances. Municipalities therefore had the opportunity to be rewarded for their improved implementation of welfare benefits. They tried to reintegrate social welfare beneficiaries into the workforce. The acceptance of the objective dispensing model resulted partly from local governments feeling that they could control the number of allowances disbursed (Rapport Meccano 2007).

According to the Ministry of Social Affairs and Employment, the introduction of the new law and the dispensing model had the effect that social security beneficiaries more often found paid work because local government was active in mediation leading to paid work. In an evaluation of the new law, it was concluded that the number of social security beneficiaries was reduced by 10% in the period 2003–2006, whereas it was estimated that without the new law this would have been 6% (SEO 2008). The reduction in social security spending amounted to 100 million Euro.

The acceptance of the objective dispensing model was partly caused by the policy outcome (reduction in the number of social security beneficiaries). Because of the success of the new law and the acceptance of the new model, a covenant was agreed upon by national government and the Association of Dutch Municipalities. The covenant stated that another 10% reduction was to be realized in the period May 2007 to the end of 2011 (in total approximately 35,000 persons).

Several times the objective dispensing model was evaluated by the Ministry of Social Affairs and Employment and some municipalities during the implementation process. The

focus of the Ministry and the research organization was to look for the improvement of the model. A private research organization supported this evaluation by investigating improvements such as the operationalization of the variables and the use of databases. Some minor changes of the model were realised.

The objective dispensing model became the central mechanism for the allocation of money to local governments. However, if municipalities did not agree with the allocated amount because of local circumstances, they could submit a request to receive additional funding. A review committee of the Dutch government assessed the requests from local governments for supplementary payments (Toetsingscommissie 2010). Requests in the period 2004–2008 were the lowest in 2006 and the highest in 2008. However, the requests for supplementary budgets amounted to less than 0.5% of the social security budget. As a result of the social security budget cuts, more and more local governments have been finding their way to the review committee (2009: 180 local governments).

#### *3.1.4. Conclusion*

The objective dispensing model has become an authoritative and effective tool in social security policy. The model is used as the mechanism to distribute all available social welfare budgets among local governments. It is accepted by the Ministry of Social Affairs and Employment, organizations that implement social security laws, municipalities and other stakeholders. The authoritativeness of the model resulted partly from the break with unpopular policies of the past. Its authoritativeness was also derived from the belief that the number of the social security beneficiaries would be controlled. In particular, local governments greatly appreciated the rewarding of proactive employment measures. Whether the model will remain authoritative is uncertain (APE 2010).

### **3.2. Case study 2: discharge of warm water**

#### *3.2.1. Background*

A new assessment system for warm water discharge was introduced on 21 June 2005. A vital part of the system was a new tool that calculates the environmental effects of warm water discharge. Industrial companies and energy companies are required to have a licence based on the Law on Pollution of Surface Water (WVO) that prohibits serious environmental effects. The assessment system provides guidance for standards relating to: (1) the intake of water, (2) the area where the discharged warm water touches the surrounding water (the mixing zone); this mixing zone cannot be greater than 25% of the cross-section of the watercourse in order to provide fish the opportunity to escape the heat plume and (3) the warming of surface water. The method for calculating the mixed zone is based on a three-dimensional simulation (3D) model, consisting of variables such as the flow rate, flow patterns and the amount of running water.

#### *3.2.2. Development of the simulation model during policy formation*

The assessment system used since 1975 was simple but not very refined. It was not permitted to discharge water warmer than 30°C. The assessment system was not based on knowledge available about the effects of discharge of salt or fresh warm water (WL Delft Hydraulics 2003). Since the consequences of warm water discharge can be fatal for fish, the new assessment system used fish as the reference organism. The policy aimed at

avoiding negative consequences for fish even in relatively unfavourable and rare circumstances.

The scientific knowledge available in several countries was used to develop a model. The specialized governmental research agency (RIZA), responsible for inland water management and water purification, was asked by the Ministry of Traffic, Waterways and Public Works to develop a new assessment system (RIZA 2004a, 2004b). Because energy companies and industrial enterprises were keenly interested in the issue, both sectors were invited to actively discuss the new system, the new simulations model and the calculation models, as well as the consequences for the authorization of permits. After literature research, the new assessment system was developed in an open communication process with industrial enterprises and the energy sector. This resulted in a new assessment system for warm water discharge as presented in the report of the Dutch national coordination commission on integral water management (CIW, Ministerie van Verkeer en Waterstaat 2004).

This system ensured (1) that the environmental effects of the heat discharge were acceptable and (2) that there were sufficient discharge opportunities. The new system was designed to assess the local effects of warm water discharge accurately. It was expected that the new system would increase opportunities for warm water discharge whilst at the same time safeguard the acceptable living conditions for fish.

The report described problems with water intake and warm water discharge as well as the new policy concerning the new assessment system, including new standards. The CIW report was clear in its goals but did not specify the tools necessary for the application of the system in practice.

Before the publication of the report, the industrial and energy sectors complained that the new system needed further specification but the minister chose to implement the new assessment system without delay. Consequently, the acceptance of the model in the industrial and energy sectors was not without hesitation. It was stated that some aspects of the model (definitions, variables and their relations) were insufficiently specified. However, the energy sector and industrial enterprises committed themselves to submit requests for a new licence based on the new model. The representatives accepted the new policy because the negativity about the old policy was stronger than the shortcomings of the new model based on scientific research.

### *3.2.3. The application of the new assessment system during policy implementation*

An evaluation of the introduction of the new assessment system (QA<sup>+</sup> 2007) showed that the licensors and the companies assessed the idea of the new system positively. For relatively simple situations in which a quick scan was applied, the implementation process went smoothly. The application process for companies to have a quick scan was not too complicated, and the assessment by licensors was also relatively easy. However, some private organizations were not informed about the new assessment system. Furthermore, it sometimes proved difficult to choose between the quick scan and the 3D simulation. Over time, the implementation process improved gradually, by more and earlier knowledge transfer, and by organizing communication between companies and the competent authorities.

For more complex requests for heat discharge, the implementation process proved difficult. Discussions arose between permit applicants and the competent authorities about the definitions of the assessment system, the framework conditions for the water



system, acceptance of the data used and the handling of the calculation tools. In one case, the results of the 3D simulation delivered by a specialized research company to the energy company did not match the results of the 3D simulation of the competent authorities. As a result, the authorities had on the one hand to deal with the permit application and on the other to evaluate the scientific value of the new 3D simulation. An additional source of irritation during implementation was that an investment of 40,000–70,000€ in 3D research by a specialized research organization that was qualified to use the new assessment system did not provide the intended clarity.

Mainly because of the problems in the licensing procedures, the ministry, industrial enterprises and energy companies decided to fine-tune the new assessment model. Under the direction of the Ministry of Traffic, Waterways and Public works and supported by governmental research agencies (WL Delft Hydraulics and RIZA), the model was improved by elaborated definitions and adapted operationalization of variables. Furthermore, a protocol for the implementation of 3D simulations was developed. This clarified details about the application of the method, and described procedural steps and the different responsibilities during each of the procedural steps (Deltares 2008). Some definitions were changed and some tools were described in more detail. The new protocol offered both the firms and the authorities support in their communication. Consequently, the course of the assessment process had a predictable outcome for both parties (WL Delft Hydraulics 2007). In sum, after the increase in implementation problems, the CIW report was elaborated with a new protocol that provided the authorities and firms clarity and predictability about the implementation process.

#### 3.2.4. Conclusion

All in all, the new assessment method for the effects of warm water discharge became authoritative – but not without problems. After several years, the decision to allow warm water discharge and to provide the licence was based completely on the new model. Whereas the model was too abstract as initially introduced, it was further refined and fine-tuned during implementation. Following criticism from industrial enterprises and energy companies, the procedural steps and the tools used during assessment were improved. As a result, the permit applicants and licensors were clear about the new policy and its implementation. After 4 years, the model was accepted by the government, industrial enterprises and energy companies. The model and the related instruments have helped to prevent negative environmental impacts but allow warm water discharge when possible.

### 3.3. Case study 3: Energy Performance Building Directive (EPBD)

#### 3.3.1. Background

Since 1 January 2008, Dutch homeowners who sell their property are required to provide an energy label to the new owner. The label indicates the energy performance of the property and therefore the environmental consequences in terms of CO<sub>2</sub> emissions. The energy performance and therefore the label are based on a model that calculates the Energy Index using variables such as the use of insulation, double glazing and the heating appliance. The label is the primary tool of government policy to obtain energy saving in the existing housing stock. In order to be effective and to limit the use of energy in households, the label should be used not only as a description of the actual energy

performance but also as a legal norm. The energy label for buildings has the status of a legal standard, but the energy label for homes does not.

### *3.3.2. Development of the EPBD during policy formation*

In the early 1990s, knowledge was developed by specialized private research organizations about the energy performance of houses and buildings. This knowledge was processed into a legal standard for the energy performance of new houses and buildings (the energy performance coefficient), which came into force on 1 January 1995. The so-called energy performance standard (EPN) offered government a tool to improve the energy performance of new houses. The method for calculating the energy performance coefficient has been the subject of extensive research. Over the years, the results of this research were programmed and discussed within the Dutch normalization institute (NEN). Within this organizational setting, private research organizations, policy officials of governments, and private organizations in the building sector cooperated to develop the professional standards to assess the energy performance.

After the introduction of the EPN for new buildings, a model was developed for the energy performance of the existing houses. This resulted in the introduction of the (not compulsory) energy performance advice (EPA) in 2003. In the same period, in 2002, the European Parliament adopted the EPBD. Since then, the EPBD has been the main legislative instrument for improving the energy performance of the building stock (European Union 2011). It took until 1 January 2008 before the EPBD came into force in the Netherlands. The method of calculating the EPBD (Energy Index) is a further development of the EPA model. The Energy Index is processed in review guidelines that are accepted as the standard for the professional groups in the building sector (ISSO 2008a, 2008b). The review guidelines are seen as a way to improve the quality of the Energy Index and thus to guarantee the quality of the energy label. The Dutch model is elaborate but in comparison with other countries not overly detailed (European Union 2011).

In addition to the development of the guidelines for the Energy Index, the Dutch normalization institute developed the Energy Diagnosis Reference (EDR) that contains standards to assess whether a specific calculation method (or software in which these methods are used) is suitable to measure houses' energy performance and to calculate the energetic and financial consequences of saving measures (Beoordelingsrichtlijn 9501). Another assessment directive contains standards for companies that deliver energy advice (Beoordelingsrichtlijn 9500). These standards relate to both the training requirements for persons and to demands made on organizations. The calculation method for the Energy Index has been carefully specified and is included in teaching materials of various professional groups in the building sector. Quality is assured by random checks of the characteristics of the energy labels awarded.

After more than a decade of scientific research, models have been developed to describe and explain first the energy performance of new homes and then the existing houses and buildings. The models have been accepted by the Ministry of Housing, private organizations in the building sector, representatives of homeowners, the social housing sector and representatives of private tenants. The urgent need for policy instruments to reduce CO<sub>2</sub> emissions supported the introduction of the models.

### 3.3.3. *The application of the EPBD during policy implementation*

By the end of 2010, 1.7 million residences had been provided with an energy label (24% of the housing stock). Decisions about the label classification of houses and buildings are completely based on the outcomes of the model. Especially as a result of the active cooperation of the housing corporations, large groups of properties have acquired an energy label. Because of the similarities of houses in the same blocks, great quantities of labels could be delivered in a relatively cheap and easy way. Compared to tenants, owner-residents rarely have an energy label (approximately 5% of owner-residents). This cannot be explained by a lack of qualified advisors in business to deliver the energy label. About 250 companies are qualified to provide it, and around 200 companies are qualified to deliver advice that includes energy conservation measures (e.g. investment costs, payback times, and environmental effects).

The acceptance of the model (Energy Index) was put to the test in practice. On the national TV news, it was reported that the results of the calculations (Energy Index and energy labels) proved to be arbitrary. A study by the Inspectorate of the Ministry of Housing in 2010 showed that 8 out of 30 calculations resulted in an Energy Index that differed substantially. In 4 of the 30 cases, the houses should have been given a different label class (VROM 2010). A study by an association of owner-residents (Vereniging Eigen Huis) came to the same conclusion. Property characteristics were often recorded erroneously (VROM 2010). Incorrect imports into the software and the incorrect execution of instructions and schemes were the most important sources of errors.

In order to meet the expectations of the inspectorate, the calculation method was improved. In actual practice this was done in the Dutch normalization institute that provides the platform to improve professional standards in the building sector. The following measures came into force on 1 January 2010: (1) definitions and extensions of tables were adjusted and (2) instructions for the calculation method were improved (for example, calculation method for the thermal peel). Furthermore, the quality of organizations and advisors was assessed which resulted in the withdrawal of the accreditation of 25% of the approved people/organizations. In addition to this, new requirements for EPBD advisors and new exams for the existing advisors were imposed.

In addition a procedure was developed that prescribed the way to handle complaints about a given label. Dissatisfied homeowners or housing cooperatives can send their grievances about their label to a complaints desk. In addition, homeowners can address disputes about the Energy Index and/or the label to the National Disputes Committee. Since 2010, no complaints have been submitted to this committee. Apparently, the model is the determining factor in the choice of energy label to classify energy performance (Geschillencommissie 2011).

At the end of 2010, the minister announced measures to ensure that energy labels were provided when homes changed hands. Without an energy label attached to the legal transfer of ownership, a house cannot be registered to the new owner in the land registry (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties 2010). Furthermore, sanctions are foreseen if the energy label is not included in sales or rental advertisements. As yet, no policy decision has been made on whether to set minimum standards for the energy performance of the existing houses.

### 3.3.4. Conclusion

The model (Energy Index) is accepted as the determining factor for the classification of house energy labels. The model has endured criticism since its introduction in 2008, but because of improvements in the model and the instructions that support its implementation, the Energy Index has become the most important policy instrument to reduce CO<sub>2</sub> emissions in houses. The policy focus has now shifted to a drive to ensure that labels are provided to the new owners when houses are sold.

## 4. The three ravens compared

Comparison of the case studies shows that none of the policy implementation arenas studied proved to be completely evidence resistant. The scientific knowledge, the models and calculation methods were not only used as starting points of new policy, but also contributed to policy implementation. However, the case studies differ in the extent to which the authority of the models was accepted. In cases 2 and 3, the models were criticized and required amendment. In other words, the ravens were not as white as we initially assumed. During the process their colours fluctuated, only gradually turning white and perhaps leaving some black feathers or a veil of grey.

The comparative analysis of the cases brings to light the following factors that influence the authoritativeness of scientific knowledge.

### 4.1. The quality of scientific knowledge

In all three cases, policymakers invested in the development of scientific knowledge over several years. The objective dispensing model, the 3D simulation and the Energy Index were the result of elaborate and solid scientific research. The various research institutes were capable of demonstrating the quality of the developed models and calculation methods. The lack of conflict over the models, despite their redistributive implications, may be explained by the fact that their scientific bases were accepted by policymakers, implementing organizations, target groups, representatives of target groups and stakeholders in the implementation process. The models contributed to the depoliticization of debates, although not to the same extent in all cases. In the warm water discharge case, the model was less well developed than in the other cases. As a result, the knowledge base of the model was questioned. The social security case shows that, although a sound scientific knowledge base is important, other factors contribute to the authoritativeness of the models too.

### 4.2. Adjustment to specific circumstances

The models and methods in each of the case studies were designed to deal with the specific characteristics of the situation in which they were to be applied. The objective dispensing model uses data relating to the regional or local situation as provided by the Central Statistics Office. Although the models aim at describing a specific regional or local situation, a need for adjustment arose when the models were initially applied in the policy implementation process. Apparently, there is a gap that is hard to bridge between generic objective calculation methods and the unique, specific situations in implementation practice. Although the models were developed to describe regional and local circumstances, they failed to do so in detail. This seems to be a fundamental barrier to the

knowledge base in public policy. In each of the cases, it proved necessary to further improve and fine-tune the model and methods during implementation.

#### ***4.3. Methods to safeguard scientific quality during implementation***

The translation of the method to its application in policy implementation practices is susceptible to errors. The data input for the Energy Index went wrong because certified consultants made more mistakes than expected. Up to 25% of the certified consultants lost their qualification as EPBD advisor because of errors in the application of the method. In the warm water discharge case also, the application of the model had to be improved. This was accomplished by making agreements on a strict research protocol for 3D modelling among the different parties involved. No matter how strong the scientific knowledge base of policy is, arrangements are needed to safeguard scientific quality during the implementation process.

#### ***4.4. Involvement of target groups and experts in knowledge production***

In none of the three case studies were the models developed completely separate from policy implementation practice. From the start, target groups, stakeholders, relevant representatives of policy actors and experts were involved in the development of the models and the methods. This may explain why the parties involved trusted the models and the policies in which these were used. Stakeholder involvement is not without risks though. It can easily result in ‘capture,’ where the results of the research are influenced by the interests of stakeholders (Wilson 1989). In the warm water discharge case, the involvement of the energy sector and industry could have impeded the effectiveness of the new system because of their interest in reducing costs and preventing the loss of profits. In this case, the risk of capture was countered by the active involvement of other stakeholders, governments and research organizations. Despite the involvement of stakeholders in ex ante knowledge production, the policies and models needed adjustment during their application, as the other factors indicate. Apparently, early stakeholder involvement is not enough to gain authoritativeness.

#### ***4.5. Opportunities for stakeholders to anticipate a new policy practice***

Because of their involvement during research activities, implementing organizations and stakeholders were well informed about the characteristics of the models and their consequences. Thus, they were able to anticipate the new policies and the models used. This enhanced their acceptance. In the objective dispensing case, municipalities could anticipate its introduction by improving their proactive employment policies in order to avoid budget cuts. In the warm water discharge case, the ministry chose not to take time to assess the consequences of the new assessment system. As a result, a legal battle emerged, jeopardizing the authoritativeness of the model. Eventually, the conflict was settled by the adjustment of the 3D model and by drafting a protocol for procedural steps that should guide its application in the implementation process.

#### ***4.6. Intermediary and mitigating arrangements***

Despite their sound knowledge base, the case studies show that the models could not gain authority during implementation on their own. Their generic characteristics did not

succeed in coping with the unique characteristics of implementation situations. Additional arrangements were necessary to bridge this gap. In the social security case, a review commission (Toetsingscommissie WWB) assessed whether local governments that did not accept the outcomes of the model were entitled to a supplementary allowance. In the warm water discharge case, legal procedures were used to discuss the 3D model. Consequently, procedural steps were described, and the role of the permit applicant as well as the licensor was defined. The calculation of the Energy Index can be checked by the homeowner or the housing corporation in a complaints procedure. These intermediary and mitigating arrangements support the acceptance of the model. If an actor is of the opinion that the model is not suitable for a specific situation, these arrangements are used to discuss the model and, if necessary, decisions are corrected. Such arrangements prevent the model needing further refinement in order to cope with every conceivable detail. Of course, there is a risk that parties will act strategically and try to systematically bypass the model. The fact that in the social security case 180 municipalities asked for an additional budget allocation in 2009 is an indication of this risk (Toetsingscommissie 2010). This risk can partly be countered if the review commission refrains from opportunistically making exceptions to avoid conflicts and only makes exceptions on the basis of substantive arguments.

#### ***4.7. Interests of the parties in the results of the model***

In the three cases, the new models might have become a serious threat to the parties involved in policy implementation, since the models reallocated gains and losses. However, in the social security case, municipalities that were going to receive less assistance funding did not resist the model. The perception that they had the opportunity to improve their employment measures and increase their incomes may explain this. In addition, many disliked the old system of social security delivery. This historical context provided a shared interest in the new model. In the warm water discharge case, the outcome of calculations using the model might result in a refusal to grant a licence. This could mean that a firm had to postpone its production process or even that it had to move. At the same time, the companies had a shared interest in changing the old system of assessment. In most cases, the new system offered the possibility to discharge more warm water than before. The Energy Index and the energy label were only to inform homeowners and did not include the obligation to invest in energy conservation measures. The moderate response of homeowners reflected the academic and nonbinding nature of the energy label. The difference between this and the other two cases is that this seemed to result in nonimplementation. It may well be that this raven will turn black when the current policy becomes obligatory.

### **5. Conclusion: lessons learned from the three ravens**

As it turns out, in policies in which the use of expertise and research affects the existing distribution of costs and benefits among relevant parties, knowledge and rationality may still matter. In the cases at hand, models based on scientific research contributed to the development of new implementation practices and to the quality of the policy. Nevertheless, it is quite a challenge to make knowledge authoritative in such an implementation setting. The cases differed in the degree to which they lived up to their status as white ravens in the implementation process. In cases 2 and 3, the models became

the subject of conflict, but even though the negative media attention in case 3 damaged support for the model, it was not altogether fatal to its authoritativeness. What generic lessons can be learned from these cases about the conditions under which knowledge can become authoritative in implementation practices?

The first lesson relates to the quality of the knowledge-generation process. The knowledge in the various cases had a solid scientific foundation; much time and expertise were invested in the development of the models. Although this was an important success condition, it was neither the only nor the decisive one. Another very important factor was the fact that stakeholders were involved in the process. They participated in discussions about the conditions under which the models were developed. In these cases, knowledge was not injected into the practice, which would have been in line with an instrumental approach to knowledge use. This participation in knowledge production, however, may be subject to certain requirements. Because of capture, interests may penetrate into the knowledge-generation process to such an extent that the research results will not bear the scrutiny of criticism (de Bruijn and ten Heuvelhof 2002). It is important that representatives of the various stakeholders participate in the process, and that the boundary work that safeguards the independence of the expertise vis-à-vis the policy process is adequately executed (see also Jasanoff 1994; Bijker, Bal, and Hendriks 2009; Gieryn 1983).

A second lesson concerns the way in which the generated knowledge is applied in the policy practice during implementation. Knowledge-based policy designs, or evidence-based policies, cannot simply be deployed in practice. Authoritativeness requires continuous interaction between knowledge producers, policymakers and implementers, and fine-tuning. The cases demonstrate three ways to adjust knowledge-based policies to specific implementation reality: fine-tuning, safeguarding the quality of knowledge during implementation, and mitigation and compensation mechanisms. In addition, policymakers should find ways to align interests (will) and knowledge. Will and knowledge do not necessarily exclude each other.

Thirdly, in looking for this alignment a major risk is that knowledge generation will dissolve in the social negotiation process between policymakers and stakeholders. Since the process of knowledge production and the application of knowledge cannot be separated, strategic behaviour by actors may drive out knowledge in both these processes. Arrangements are needed to mitigate strategic behaviour and safeguard the quality of scientific knowledge generation. The absence of conflict in the three cases might indicate that actors succeeded in manoeuvring unwanted evidence out of the arena. Although our research observed the risk of this type of behaviour, our observations regarding the development of the models, the impacts of these models on the policies and the way the processes evolved indicated that these tendencies were limited and did not compromise the objective, scientific nature of the models.

This article has identified several mechanisms that affect the authoritativeness of knowledge in three critical cases of policy implementation in which scientifically developed models distributed advantages and disadvantages. Our knowledge about these mechanisms and the pitfalls involved remain limited and call for further research. Firstly, further research is needed to establish whether or not these mechanisms manifest themselves in the same way in other policy settings. Secondly, further research might identify other mechanisms that may be at play. Thirdly, research is needed to learn more about the relative importance of the factors: are these factors sufficient or necessary conditions, and can one factor be compensated by another?

Practically, the case studies demonstrate that scientists and policy implementers are not empty handed when it comes to furthering the use of scientific knowledge in the practice of public policymaking and implementation. They have a repertoire of options that range from involving stakeholders in the knowledge-production process, to adapting knowledge basis in between, to refining it, to integrating it with tacit expertise, commons knowledge and emerging new insights (learning), to installing compensating mechanisms and prudent implication.

### Acknowledgements

We would like to thank the researchers of the group Comparative Public Service Innovation (CPSI) at the Department of Public Administration, Faculty of Social Sciences, Erasmus University Rotterdam, for their constructive comments on an earlier version of this article.

### References

- Aarts De Jong Wilms Goudriaan Public Economics bv (APE). 2003. "A new allocation model for the Law on Social Security and Work." In *Minutes Second Chamber of the Dutch Parliament, 2002–2003*, 28.870, nr. 13.
- Aarts De Jong Wilms Goudriaan Public Economics bv (APE). 2006. "The Improvement Potential of the Objective Allocation Model." In *Minutes of the Second Chamber of the Dutch Parliament, 2005–2006*, 30.545, nr. 3.
- Aarts De Jong Wilms Goudriaan Public Economics bv (APE). 2010. "Macrobudget Inkomensdeel WWB 2010 [Macrobudget Incomes share Law on Work and Social Assistance 2010]." In *Council Municipality of The Hague, BSW 2010–2011, RIS175780*.
- Banks, G. 2009. *Evidence-based Policy Making: What Is It? How do We Get it?* Accessed June 4, 2010. <http://www.pc.gov.au/speeches/cs20090204>.
- Berkhout, F. 2010. "Reconstructing Boundaries and Reason in the Climate Debate." *Global Environmental Change* 20 (4): 565–569. doi:10.1016/j.gloenvcha.2010.07.006.
- Bijker, W. E., R. Bal, and R. Hendriks. 2009. *The Paradox of Scientific Authority. The Role of Scientific Advice in Democracies*. Cambridge, MA: The MIT Press.
- Blatter, J., and M. Haverland. 2012. *Designing Case Studies. Explanatory Approaches in Small-n Research*. Houndsmills Basingstoke: Palgrave Macmillan.
- Collins, H., and R. Evans. 2007. *Rethinking Expertise*. Chicago: The University of Chicago Press.
- Davies, H., S. Nutley, and P. Smith, eds. 2000. *What Works? Evidence-based Policy and Practice in Public Services*. Bristol: The Policy Press.
- de Bruijn, J. A., and M. Leijten. 2010. "Megaprojects and Contested Information." In *Decision-making on Mega Projects*, edited by H. Priemus, B. Flyvbjerg, and B. van Wee, 84–101. Cheltenham: Edward Elgar.
- de Bruijn, H., and E. ten Heuvelhof. 2002. "Policy Analysis and Decision Making in a Network: How to Improve the Quality of Analysis and the Impact on Decision Making." *Impact Assessment and Project Appraisal* 20 (4): 232–242. doi:10.3152/147154602781766627.
- Deltares. 2008. *Richtlijnen voor de rapportage van 3D-modelsimulaties* [Guidelines for Reporting on 3D-simulation Models]. Lelystad: Ministry of Transport and Water Management, RIZA.
- European Union. 2011. *Implementing the Energy Performance of Buildings Directive, Country Reports 2010*. Brussels: European Union.
- Flyvbjerg, B. 2006. "Five Misunderstandings about Case-study Research." *Qualitative Inquiry* 12 (2): 219–245. doi:10.1177/1077800405284363.
- Flyvbjerg, B. 2008. "Public Planning of Mega-projects: Overestimation of Demand and Underestimation of Costs." In *Decision-making on Mega Projects*, edited by H. Priemus, B. Flyvbjerg, and B. van Wee, 120–144. Cheltenham: Edward Elgar.
- George, A., and A. Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Geschillencommissie. 2011. *Jaarverslag 2010* [Year Report 2010]. The Hague: Geschillencommissie.
- Gibbons, M., C. Limoges, H. Nowotny, S. Scott, and M. Trow. 1994. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.



- Gieryn, T. F. 1983. "Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists." *American Sociological Review* 48 (6): 781–795. doi:10.2307/2095325.
- Head, B. W. 2008. "Three Lenses of Evidence-Based Policy." *Australian Journal of Public Administration* 67 (1): 1–11. doi:10.1111/j.1467-8500.2007.00564.x.
- Hess, C., and E. Ostrom, eds. 2006. *Understanding Knowledge as a Commons. From Theory to Practice*. Cambridge, MA: MIT Press.
- Hoppe, R. 1999. "Policy Analysis, Science and Politics, from 'Speaking Truth to Power' to 'Making Sense Together.'" *Science and Public Policy* 26 (3): 201–210. doi:10.3152/147154399781782482.
- in 't Veld, R. J. ed. 2000. *Willingly and Knowingly: The Roles of Knowledge about Nature and the Environment in Policy Processes*. Utrecht: Lemma.
- ISSO. 2008a. *Handleiding energieprestatie advies utiliteitsbouw. (75.1, 75.2)* [Manual Energy Performance Advice Non-residential Buildings]. Rotterdam: ISSO.
- ISSO. 2008b. *Handleiding energieprestatie advies woningen (82.1, 82.2)* [Manual Energy Performance Advice Homes]. Rotterdam: ISSO.
- Jasanoff, S. 1994. *The Fifth Branch. Science Advisers as Policy Makers*. Cambridge, MA/London: Harvard University Press.
- Koppenjan, J. F. M., and E. H. Klijn. 2004. *Managing Uncertainties in Networks*. London: Routledge.
- Levy, J. S. 2008. "Case Studies: Types, Designs, and Logics of Inference." *Conflict Management and Peace Science* 25 (1): 1–18. doi:10.1080/07388940701860318.
- Mazmanian, D. A., and P. A. Sabatier. 1981. *Effective Policy Implementation*. Lexington, MA: Lexington Books.
- Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. 2010. "Directoraat Wonen, Werken en Integratie, Sanctionering Energielabel Gebouwen [Agency Living, Work and Integration: Sanctioning Energylabel Buildings]." In *Handelingen Tweede Kamer, vergaderjaar 2011/2012, kamerstuk 30.196 nr. 124*.
- Ministerie van Verkeer en Waterstaat. 2004. *CIW-beoordelingssystematiek warmtelozing* [Assessment Method Thermal Discharges of the Coordination Commission on Integral Water Management (CIW)]. The Hague: Ministry of Transport and Water Management.
- Noordergraaf, M. 2000. *Attention! Work and Behaviour of Public Managers Amidst Ambiguity*. Delft: Eburon.
- Nowotny, H., P. Scott, and M. Gibbons. 2001. *Rethinking Science. Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity Press.
- O'Toole, L. J. 1988. "Strategies for Intergovernmental Management: Implementing Programs in Interorganizational Networks." *International Journal of Public Administration* 11 (4): 417–441. doi:10.1080/01900698808524596.
- Patton, M. 1997. *Utilization-focused Evaluation*. 3rd ed. London: Sage.
- Patton, M. Q. 2001. *Qualitative Research and Evaluation Methods*. 2nd ed. Thousand Oaks, CA: Sage.
- Pawson, R. 2006. *Evidence-based Policy: A Realist Perspective*. London: Sage.
- Polanyi, M. 2009. *The Tacit Dimension*. Chicago: The University of Chicago Press.
- QA<sup>+</sup>. 2007. *Evaluatie Vergunningverlening beoordelingsystematiek Warmtelozing* [Evaluation Assessment Method of the Licensing System Thermal Discharges]. Leiden: Questions, Answers and More.
- Radin, B. 2002. *Beyond Machiavelli—Policy Analysis Comes of Age*. Washington, DC: Georgetown University 1101 Press.
- Rapport Meccano. 2007. *Werkt de WWB? Resultaten van de ontwikkeling van nieuwe verhoudingen tussen rijk en gemeenten* [Does the Law on Work and Social Assistance Work?]. Utrecht: Meccano.
- RIZA. 2004a. *Effecten van koelwater op het zoete aquatische milieu* [Impacts of Thermal Discharges on Inland Water Quality]. Lelystad: Ministry of Transport and Water Management, RIZA.
- RIZA. 2004b. *Effecten van koelwater op het zoute aquatische milieu* [Impacts of Thermal Discharges on Salt Water Environments]. Lelystad: Ministry of Transport and Water Management, RIZA.

- Sabatier, P. A., and H. C. Jenkins-Smith. 1993. *Policy Change and Learning. An Advocacy Coalition Approach*. Boulder, CO: Westview Press.
- SEO. 2008. *Evaluatie verdeelmodel Wet Werk en Bijstand* [Evaluation of the Allocation Model of the Law on Social Security and Work]. The Hague: SEO.
- SEOR. 2005. *Inventarisatie verbetermogelijkheden objectief verdeelmodel WWB* [Inventory Improvement Potential of the Objective Allocation of the Law on Work and Social Assistance]. Rotterdam: Erasmus University Rotterdam.
- Solesbury, W. 2001. *Evidence Based Policy: Whence it Came and Where it's Going* (Working paper 1). London: University of London.
- Toetsingscommissie. 2010. *Jaarverslag 2009* [Year Report 2009]. The Hague: Ministry of Social Affairs.
- VROM, 2010. *Betrouwbaarheid van energielabels bij woningen, herhalingsonderzoek* [Reliability of Energy Labels for Homes]. Publication number: VI-2010-01. The Hague: Ministry of Housing, Physical Planning and Environment.
- Weiss, C. 1977. *Using Social Research in Public Policy Making*. Lexington, MA: Lexington-Health.
- Wildavsky, A. 1979. *Speaking Truth to Power: The Art and Craft of Policy Analysis*. Boston: Little, Brown.
- Wildavsky, A. 1995. *But Is It True. A Citizen's Guide to Environmental Health and Safety Issues*. Cambridge/London: Harvard University Press.
- Wilson, J. Q. 1989. *Bureaucracy. What Government Agencies Do and Why They Do It*. New York: Basic Books.
- WL Delft Hydraulics. 2003. *Warmtelozing in oppervlaktewater en uitwisseling met de atmosfeer* [Thermal Discharges to Surface Waters and Exchanges with Atmosphere]. Delft: WL Delft Hydraulics.
- WL Delft Hydraulics. 2007. *Implementeren van verbeteringen in het landelijk temperatuurmodel* [Implementing Improves in the National Warmth Measuring Model]. Delft: WL Delft Hydraulics.
- Yin, R. J. 2009. *Case Study Research: Design and Methods*. 4th ed. Thousand Oaks, CA: SAGE.