Playing a role - but which one? : how public service motivation and professionalism affect decision-making in dilemma situations

Schott, C.

Chapter 6

QUANTITATIVE METHODS AND TECHNIQUES

This chapter elaborates on the research design of the quantitative part of this study. I first explain how the questionnaire was developed and which measurement instruments were used to measure to concepts under study while paying close to their reliability and validity (6.1). After that I discuss how the data were collected, who the respondents were, and which statistical techniques were used to analyse the data (6.2). Finally, I describe how I have handled missing values (6.3) and discuss the limitations of the quantitative research design (6.4).

6.1 Developing a questionnaire and assessing the reliability and validity of the measurement instruments

The topic of this section is the structure of the questionnaire and the various stages of the pilot study. I also elaborate on each instrument used to measure the concepts I was interested in this study. Information about the reliability and construct validity of the separate measurement instruments is also provided.

6.1.1 Structure and stages of development

As suggested by Van der Velde et al. (2004), I structured the questionnaire in such a way that clusters of items were introduced in a logical sequence. The questionnaire consisted of nine parts – i.e. clusters of items – and a cover letter. In this letter the goal of the research project was briefly described, anonymity and confidentiality were guaranteed, the independent position of the research team was highlighted, some instructions on how to answer the questions were given, and the researchers' contact details were provided in case respondents wanted additional information or had questions.

The first part of the questionnaire contained questions about the demographical background of the respondents. Second, the professional role identity of veterinary inspectors was measured. I decided to ask questions related to this concept at an early stage, because these directly referred to the work of veterinary inspectors and so could trigger respondents' attention to and interest in this study. The third part was about PSM. Part four related to the organizational identification of veterinary inspectors; part five to proactive behaviour. The data collected in parts six and seven on organizational socialization are not part of this dissertation, but will be included in a related thesis by Daphne van Kleef. Finally, decision-making was measured in part eight, by confronting respondents with hypothetical dilemmas and asking them what they would do in such a situation. At the end of the questionnaire respondents were given the opportunity to write down any remarks they wanted to make about the questionnaire (which will be discussed in Subsection 6.2.2), and to provide their contact details so that I could send them the results of the study.

The questionnaire was developed over several stages, in order to maximize validity and reliability. First, I conducted an intensive literature research on existing measurement instruments and – if nothing suitable was available – on ways how to construct items of the questionnaire myself. This search resulted in a first draft of the questionnaire containing all parts described above. Second, this draft was discussed with Jeroen Maesschalck (an expert in the field of designing and developing surveys, including questions on ethical decision-making in dilemma situations) and his research team, in order to tackle methodological problems in advance (pilot test 1). After the questionnaire had been revised it was presented to four individuals who were familiar with the working context of our sample – veterinary inspectors who work in management or policy advisory functions – and a former Belgian veterinary inspector (pilot test 2). I asked them to 'think aloud' when they completed the questionnaire. This is an appropriate technique to investigate if the questions are interpreted correctly and if adequate terminology is used (Van der Velde et al., 2004). In order to avoid problems related to survey translations across different languages the quality of the translation was assessed via a carefully monitored translation process.

6.1.2 Measurement instruments

In this subsection I describe each of the measurement instruments and controls used in this study while paying attention to their reliability and validity. Table 6 provides a summary of descriptive statistics of all instruments. It includes Cronbach's α for the measurement instruments – used as an estimate of reliability –, the standard deviations (SD), and the number of items used to measure the overarching concepts and their separate dimensions. Most items were measured on a five-point Likert scale (ranging from 'agree' to 'disagree'). For the concept of work-related tensions a 10-point Likert scale was used. In the Appendix complete lists of all items used in this study can be found (Tables A4a, A4b, A4c). The original Dutch formulation of the items can be obtained on request.

6.1.2.1 Public service motivation

In 2013, Kim et al. published a modified version of Perry's original measurement scale (1996), which is shorter and cross-validated among 12 countries. In this study I measured PSM with a slightly adapted version of the Kim et al. scale. As the Netherlands were included in this international study, the Dutch version of the PSM scale could be obtained from the researchers themselves. In pilot test 2, however, one item of the dimension 'self-sacrifice' was found to cause confusion among the respondents (PSM_SS2: I believe in putting civil duty before self). This item was excluded from the follow-up survey. Next to this, I added an

open question to the PSM scale asking individuals to indicate by just one single word with what they associate the 'public interest' in their roles of A) citizen, B) veterinarian (with an academic education) and, C) veterinary inspector.

Neither of the confirmatory factor analyse (CFA) fit indices on our measurement of PSM met the required threshold. (See Subsection 6.2.3.1 for more information on CFA: an analysis to adress construct validity.) For this reason, two items (PSM_APS_4: It is important to me to contribute to the common good, and PSM_COM_ 2: I emphasize with others who face difficulties) with low factor loadings were excluded from further analysis. As a result, both fit indices increased, which indicates that the adjusted PSM model can be considered more suitable (Table 4). The CFI value exceeded the cut-off score of .95, and the RMSEA was below .08 (RMSEA = 0.076 < .08; CFI = .954 > .95), which suggests that the modified PSM measurement instruments had good construct validity.

Cronbach's α was acceptable for both the overarching concept of PSM (α = .83) and its separate dimensions 'attraction to public service', 'commitment to public values', 'compassion' and 'self-sacrifice'. Only the dimension 'commitment to public values' scored low (α = .56). Since Cronbach's α reliability coefficient depends on the number of items by which a concept is measured (Dooley, 2001) and the dimension 'commitment to public values' is addressed by only three items, I retained the dimension.

Table 4 Summary of confirmatory factor analysis

	CFI	**RMSEA**
PSM (original model)	.923	.121
PSM (modified model)	.954	.076
Cut-off criteria	>.95	<.08

6.1.2.2 Professionalism

The concept of professionalism was measured in two ways: as the extent to which respondents identify with their profession or professional role (professional identification), and as the interpretations individuals bring to their professional role (professional role identity). Both measurement instruments are discussed in detail in the following subsections.

Professional role identity

To my knowledge, there is no measurement instrument which records the interpretations that individuals bring to their professional role of veterinary inspectors. To this end, I developed a new instrument. Following research in which identity theory was used (e.g., Reitzes &

Burke, 1980; Stets & Burk, 2005a), and using insights from Cheung (2008), who developed a scale to measure teachers' professional identity by asking respondents to indicate how committed they were to their professional practices, I developed a bipolar scale of veterinary inspectors' professional role identity. This 11-item scale measured four different aspects of the professional role of veterinary inspectors, derived from the findings of the qualitative study: commitment to economic interest, commitment to animal welfare, commitment to public health, and strict rule enforcement (section 5.3). Each dimension was measured by a varying number of items, from two to three. The dimension *strict rule enforcement* was used to assess if veterinary inspectors were willing to enforce rules and regulations in each and every situation, or if they considered generating rule compliance a superior method. The remaining dimensions measured which interests veterinary inspectors find important to promote in their work: *economic interest, animal welfare,* or *public health.* Respondents were asked to think of themselves in their position as veterinary inspectors and prompted to identify where they would place themselves between the bipolar statements. In other words, they had to indicate to what degree they perceived the four dimensions as important aspects of their professional role. Some examples of the items: 'Strict enforcement of rules is the only way to reach your goals (strict rule enforcement)' or 'It is important that veterinary inspectors consider the economic interest of the meat processing industry (commitment to economic interest)'.

In Table 5 the results of the principal component analysis (PCA) are summarized. (See Subsection 6.2.3.1 for more information on PCA: an analysis to adress construct validity.) Before the results of the PCA can be analysed to assess the construct validity of professional role identity, it is important to check whether the Kaiser-Meyer-Oklin (KMO) test and Bartlett's test of sphericity meet the required criteria (Field, 2009). The KMO test indicates whether the sample is adequate for the analysis to be carried out. In this study, the measure of sampling adequacy exceeded the required cut-off criterion of .5 (KMO test = .661). Bartlett's test was significant (.000), which means that this matrix is proportional to an identity matrix.

The results of the PCA indicated that the factors approximately correspond to the expected aspects of professional role identity (commitment to economic interest, commitment to animal welfare, commitment to public health, and rule enforcement). All factor loadings were above .4: they ranked between .439 and .805 and the Eigenvalues were greater than 1 which – according to the Kaiser's Criterion (1960) – justifies the retention of all four factors.

The reliability scores for the separate dimensions of professional role identity were rather low (see Table 6), especially for 'commitment to economic interest'. This deserves

closer attention. Further investigation of the two items "If rule enforcement implies serious financial damage to the individual I have to inspect, I find it difficult to enforce rules (ECO1)" and "Sometimes, I deviate from the rules in order to reduce the economic damage to the individual I have to inspect (ECO2)" showed that together they form a cumulative scale instead of a scale of correlated items. In other words, finding it difficult to enforce rules if this triggers serious financial damage to inspectees was a precondition for deviating from the rules. Embretson and Reise (2000) point out that cumulative scales cannot be assessed by Cronbach's α, but should be analysed via nonparametric item-response theory for polychromous items. The scale of the dimension 'commitment to economic interest' had a homogeneity of H = 0.34, which according to Van Schuurman (2003) is acceptable. For this reason, we combined ECO1 and ECO2 and used their sum score as instrument to measure the professional role dimension 'commitment to the economic interest'.

Taking into account that Cronbach's α depends on the number of items used to measure a concept (Dooley, 2001) and that the number of items used to measure the different dimensions of professional role identity ranged only from 2 to 4, I considered the Cronbach's α values for the other dimensions of professional role identity – ranging from .61 to.66 –acceptable.

Table 5 Results of principal component analysis

	Factor			
	1	**2**	**3**	**4**
Eco1	-.476	-.041	.036	.628
Eco2	-.514	-.039	-.078	.606
AW1	-.307	.318	.439	-.017
AW2	-.098	.393	.703	-.190
AW3	-.262	.551	.534	.106
PH1	-.094	.758	-.319	.046
PH2	.030	.540	-.569	.092
PH3	-.047	.549	-.383	-.230
Enfore1	.807	.039	.112	.126
Enforce2	.749	.148	.107	.351
Enforce3	.885	.205	.135	.271
Eigenvalue	**2.67**	**1.78**	**1.59**	**1.09**

Rotation: Varimax

Professional identification

Following many other scholars (e.g., Hekman et al., 2009; Loi et al., 2004), professional identification is assessed via the rephrased organizational identification instrument developed by Mael and Asforth (1992). The rephrased instrument consists of six items. I shortened the instrument because professional identification was not one of my main points of interest. The remaining three items were "I am very interested in what others think about the profession of veterinary inspector", "When I talk about veterinary inspectors, I usually say 'we' rather than 'they'", and "If a story in the media criticized veterinary inspectors, I would feel embarrassed".

The results of the reliability analysis were interesting. Even though the scale has been validated numerous times, Cronbach's α was very low for three items (α = .32) and remained low if item 2 ("When I talk about veterinary inspectors, I usually say 'we' rather than 'they'") was deleted (α = .48). These results raise questions about the appropriateness of the professional identification scale regarding the profession of veterinary inspector. Because professional role identification was not one of the key constructs in our study, and the low α might have been the result of the small number of items used to measure professional identification, I did retain the instrument for subsequent analyses. Nevertheless it is important to bear in mind that an instrument with a low Cronbach's α increases the risk of biased results.

Table 6 Descriptive statistics for the measurement instruments

Construct	No of items	Cronbach's α	Mean	SD	No of valid observations of N=258
PSM	12	.83	3.78	.40	258
Attraction to public service	3	.58	3.84	.48	258
Commitment to public values	3	.56	3.81	.40	258
Compassion	3	.60	3.37	.51	258
Self-sacrifice	3	.62	3.37	.61	258
Professional role identity	11	.64	3.26	.38	
Commitment to economic interest	2	.48[1]	2.46	.67	258
Commitment to animal welfare	3	.60	3.32	.66	249
Commitment to public health	3	.61	3.60	.62	254
Strict rule enforcement	3	.66	2.33	.66	255
Commitment to the inspectee	2	.61	2.96	.81	256
Professional identification	3	.48	3.36	.77	258
Work-related tensions	1		5.68	2.08	219

6.1.2.3 Commitment to the inspectee

The construct 'commitment to the inspectee' was measured by a 2-item instrument on the basis of the results of the interviews discussed in section 5.6. The instrument measured the degree to which veterinary inspectors commiserate with their inspectees and the stress they experience if they have to make a decision that implies adverse consequences for the inspectee. The items were "I find it difficult to act as a strict enforcer of rules and regulations if I know that the people I am inspecting have done their best to improve things" and "I find it difficult act as an strict enforcer of rules and regulations if I know the person I am inspecting personally". Cronbach's α was acceptable (α = .61).

6.1.2.4 Work-related tension

The instrument used to measure work-related tensions was based on a 1-item measure used by Lindquist and Whitehead (1986), who measured work-related stress by asking "How stressful do you consider your job to be". I re-formulated the question in the following way: "Please indicate to what degree you experience tensions in your work as veterinary inspector". A score of '1' meant 'no work-related tensions' at all; a score of '10' indicated 'many work-related tensions'.

6.1.2.5 Decision-making in dilemma situations

Following a large body of quantitative research on ethical decision-making (e.g., Maesschalck, 2004, Loe et al., 2000), I measured decision-making by confronting respondents with three hypothetical dilemmas – or 'vignettes' – and asking them how they would act in these situations and why. These dilemmas were developed by myself on the basis of the results of the qualitative study described in section 5.1. Thus, the dilemmas were highly realistic and presented real-life problems frequently encountered by veterinary inspectors. The aim was to develop dilemmas in which the three core values of veterinary inspectors – *public health, animal health/welfare, and economic interest* - were in conflict which each other. For all three dilemmas respondents could choose between different response categories that reflected the dominant types of decisions – or coping strategies – I had derived from the interviews (5.2). The answers respondents could choose from fall into three categories: 1) safeguarding one of the values that were in conflict (to apply the coping strategy 'biasing'), 2) an interim solution (applying the coping strategy 'avoidance'), or 3) "differently, namely…". The dilemmas are presented below. In Dilemma 1, there was a conflict between economic interest and public health; in Dilemma 2, between economic interest and animal welfare; in Dilemma 3, between public health and animal welfare.

1 See section 6.1.2.2 for the reason why I retained this concept in spite of its low Cronbach's α.

In line with research on ethical decision-making (Maesschalck, 2004), I asked respondents for a number of considerations to indicate on a 5-point Likert scale how important they were for their decision-making. These considerations had been formulated on the basis of the results of the qualitative analysis described in section 5.6, and are therefore of an exploratory nature. Nevertheless, most of them fit well into the typology of nine 'ethical standards' or 'ethical climates' developed by Victor and Cullen (1988) and adapted by Maesschalck (2004). (For more information about this typology see section 5.6.) For example, "I have decided to make this decision because this is what the rules and regulations say" is associated with the ethical identity standard 'organizational rules', "Because it was my colleague who overslept" can be linked to the standard 'team interest', "Because I want to avoid any risk to animal welfare/health" fits well in 'stakeholder orientation'. The question measuring veterinary inspectors' considerations in decision-making was relevant, because it double-checked why individuals made a certain decision. In a final step the respondents were also asked to rank their three most important considerations. Asking individuals to be clear about their considerations was intended to increase the internal validity of this study. By investigating the considerations I learned whether decision-making in dilemma situations could have been caused by alternative independent variables, variables I did not include in the theoretical model in the first place. Besides, it enabled me to control whether the kind of decisions individuals made indeed matched the considerations that are expected to underlie certain coping strategies (for more information about coping strategies see section 2.2).

According to Wimbush et al. (1997), using specific dilemmas in survey research increases the likelihood of socially desirable answers, because respondents might suspect their employer of using the survey to investigate how they behave in conflict situations. However, because I expected that realistic scenarios would make it easier for veterinary inspectors to put themselves into the dilemma situations, and so give me higher-quality information than I could get from general dilemma scenarios, I decided to stick to realistic and specific vignettes. I hoped to counter the threat of socially desirable answers by strongly emphasizing in the cover letter that anonymity and confidentiality were guaranteed. The original Dutch text of the dilemmas can be obtained from the author on request.

Dilemma 1: Economic interest versus public health

An employee of an abattoir calls you in the afternoon about a post-mortem inspection. It turns out that one of your colleagues overslept this morning. The stable hand has already started the slaughtering, without waiting for a veterinary inspector to arrive for the ante-mortem inspection. You are a little surprised because usually this abattoir sticks to the rules. By the time you arrive at the abattoir, 15 heads of cattle have been slaughtered without ante-mortem inspection. What are you going to do?

1) I certify the cattle.
2) I disqualify the cattle.
3) I defer the decision until I have talked to my supervisor.
4) Something else, namely….

Dilemma 2: Economic interest versus animal welfare

In a poultry abattoir where you have to do an inspection an old-fashioned and animal-unfriendly machine is still in use, unloading all chickens at the same time. Official documents state that animals have to be kept in a horizontal position when they are unloaded. However, at this abattoir all animals are continually falling on top of each other. This increases the risk of injury and suffering. You are quite sure that the large number of broken wings is caused by the old-fashioned machine, and not by something else. You want to do something about this, but you know that your colleagues have not enforced the rules consistently in the past. Stopping the production process implies serious financial damage. What are you going to do?

1) I do not do anything.
2) I defer the decision until I have talked to my supervisor.
3) I make a written report.
4) I stop the production process.
5) Something else, namely…….

Dilemma 3: Public health versus animal welfare

At an abattoir a cow is unloaded. You suspect that the food chain information (FCI) about the animal is not reported correctly. Next to this, it turns out that one of the cow's legs is fractured. The truck driver explains that the fracture must have happened on the way to the abattoir, but it is not clear by just looking at the animal if this is true. What are you going to do?

1) I allow the cow to be shot and slaughtered.
2) I order the cow to be shot and disqualify it.
3) I order the cow to be shot and slaughtered and try to predate the fracture so that I can act if necessary.
4) I order the cow to be shot and slaughtered and defer my decision until I receive the additional food chain information I require.
5) I defer the decision until I have talked to my supervisor
6) Something else, namely ……

The following considerations were presented to the veterinary inspectors after they had decided what they were going to do. (Note that not all considerations apply to all dilemmas.)

I have decided to make this decision because:
a) this is what the rules and regulations say.
b) otherwise I will have trouble enforcing rules and regulation at this company in the future.
c) I want to avoid a negative working atmosphere if I can
d) by talking to my supervisor I play safe.
e) I want to avoid any risk to animal welfare/heath (not applicable in Dilemma 1).
f) I want to avoid any risk to public health (not applicable in Dilemma 2).
g) I want to limit the financial damage (not applicable in Dilemma 3).
h) It was my colleague who overslept (only applicable in Dilemma 1).
i) I have difficulty throwing away 'good meat' (only applicable in Dilemma 1).
j) the inspectee is usually a good guy (not applicable in Dilemma 3).
k) there has been no strict rule enforcement in the past (only applicable in Dilemma 2).
l) of something else, namely……

6.1.2.6 Controls

Because demographic characteristics and contextual circumstances can bias research findings, a number of commonly used and case-specific controls were included in the questionnaire. This enables us to account for the relationships between two or more variables that may be affected by one or more additional variables. In Table 7 (6.2.2), the descriptive statistics of the control variables of the respondents are summarized.

Following common practice (Pandey & Stazyk, 2008), the following three socio-demographic characteristics were included as control variables: *position, age*, and *gender*.

Educational level was not included because an academic degree in veterinary medicine is a requirement for becoming a veterinary inspector. Next to this, I included – for theoretical reasons or because of findings in the qualitative analysis – the additional control variables years of employment, type of employment contract (practitioner or 'regular' veterinary inspector), team (living animal, abattoirs, and import), and additional employment contract (yes or no). *Years of employment* were added because only those veterinary inspectors who entered employment after 2010 had had a very intensive introductory training focusing on strict enforcement of rules and regulations. It is realistic to be alert to the fact that this training might have affected the way veterinary inspectors interpret their professional role, the decisions they make, and ultimately the relationship between these two concepts. The *type of employment contract* is also relevant in this context, because practitioners are not as well-integrated in the organization. They work fewer hours than regular veterinary inspectors (mean = 13.3 hours as opposed to 34[2]) and do not always participate in team meetings. This weaker organizational tie might have influenced decision-making in dilemma situations and – as partly suggested by the interview results – the way inspectors perceived their professional role. The same applies to the control variable *additional employment contract.* It is reasonable to assume that veterinary inspectors who also work as veterinarians will more easily make decisions favourable for farmers, because it is the farmers who provide their income in their role of veterinarian.

Finally, *proactive personality* was included because in sociological literature this concept is commonly associated with successful adoption of organizational values (Ashford & Black, 1995). Respondents with a proactive personality, therefore, might enforce rules and regulations in a strict manner because this is what the NVWA stands for. The concept of proactive personality was measured on a 4-item scale developed by Parker and Sprigg (1999), which is based on Bateman and Crant's 17-item Proactive Personality Scale (1993). Examples of items we used in this study are "I excel at identifying opportunities" and "I am always looking for better ways to do things".

2 The mean number of working hours is based on the veterinary inspectors who responded, not the entire population.

6.2 Data collection and analysis

In this section the data collection procedure is described in detail, as well as the representativeness of the sample and the participants' remarks about the questionnaire. I also discuss the different statistical techniques used to analyse the survey data.

6.2.1 Method: survey research

The *surveymonkey software* was used to administer and distribute the questionnaire. All respondents received a link via their NVWA email address. (All NVWA employees have their own email account so that the organization can communicate with them.) An online questionnaire was preferred over a paper-and-pencil questionnaire to collect data for the survey research because, participants in the pilot study indicated that they found online questionnaires most convenient (for more information about the pilot study, see subsection 6.2.1). Other advantages of online surveys are the reduced chance of errors during data entry and the possibility to 'force' individuals to answer certain questions, reducing the percentage of missing data.

Given the small population we did not draw a sample. The web-based questionnaire was distributed to the entire population: all veterinary inspectors working for the Dutch Food and Product Safety Authority, Veterinary & Import division. In order to ensure a high response rate the questionnaire was accompanied by a short cover letter signed by the division head, which emphasized the independent and autonomous character of this study and the importance of participating. The questionnaire was distributed in June 2013. At intervals of about two weeks two reminders were sent.

6.2.2 Respondents, response rate and respondents' remarks about the questionnaire

In total, 403 employees were addressed, 269 of whom returned the questionnaire (67 %). Respondents who completed less than 30% of the questions were excluded from the analysis, which yielded a total response of 64% (N=258). Of the respondents, 63% were male and 37 % female. More than 50% of the respondents held the position of veterinary inspector without managerial responsibilities; an additional 26 % held a position as company inspector, which included coordinating responsibilities in one or more companies, and 16 % were senior veterinary inspectors, whose task is to support colleagues with their substantial expert knowledge of a specific domain. The largest age group was between 55 and 65 years old (36%), and had worked at the NVWA for between five and ten years (40%). This large group of older employees and the frequent five to ten years' duration of employment support the

assumption that many veterinarians have a 'second career' at the NVWA, after working as veterinarians in private practices for many years. Only 10 % of the sample were younger than 35 years. Sixty-five per cent had an employment contract as a 'regular' veterinary inspector, 33% worked as practitioners on call, and 32% of the sample were also employed as veterinarians in private practice. Most of the employees with an additional employment contract were practitioners. However – even though not appreciated by the NVWA – there were also eight 'regular' veterinary inspectors who worked in private practices. The demographic characteristics of the respondents are summarized in Table 7. This table also provides information about the demographic characteristics of the entire population of all veterinary inspectors working at the NVWA, and the differences between these two groups: the difference between the sample and the entire population and the results of the *chi*-square test testing for the representativeness of the sample. The reason why these particular demographic characteristics are included in this study has been explained in Subsection 6.1.3

Most of the results of the *chi*-square test were non-significant implying that the sample generalizes well to the population with respect to age, type of employment contact, and team. However, we also need to be aware of significant findings of the *chi*-square test indicating that the sample does not well represent the population with regard to gender and position. Nevertheless, because the differences with regard to position could be explained by the fact that the list of those included in the sample was more recent (February 2014) than that of the entire population of Dutch Veterinary inspectors (summer 2013), the sample can be considered highly representative. The definition of the position of senior veterinary inspector had changed between summer 2013 and February 2014, which explains why there were more senior veterinary inspectors in the sample than in the population. Gender differences proved the only limitation to the representativeness of the sample. In the population, there were more women than men (56:44), but in the sample it was the other way around (37:63).

Some respondents formulated specific remarks about the questionnaire in line with the invitation: 'for more information or questions please do not hesitate to contact [name researchers]'. One person called in order to ask if the questionnaire could be sent to a different email address. Another telephoned because he had difficulties opening the link. Most often by far respondents used the open-ended question at the end of the questionnaire to formulate general remarks about the questionnaire (n = 82). These remarks can be differentiated into six topics: 1) phrasing of the items, 2) aspects missing from the questionnaire, 3) appropriateness of the dilemmas, 4) length of the questionnaire, 5) positive feedback, and 6) other remarks. (See also Table 8). I will briefly discuss each of these topics.

Table 7 Descriptive statistics of sample and population

	Sample Total N=258		Population[2] Total N=417		Difference		*Chi*-Square	*P* Value
	N	%	N	%	N	%		
Position								
Veterinary inspector	141	55	239	57	89	2		
Veterinary inspector and company inspector[3]	68	26	148	36	80	14		
Senior veterinary inspector[4]	42	16	30	7	12	9		
Other[5]	7	3					102.04	.000
Age								
<35	25	10	39	10	14	0		
≥35 and <45	53	21	74	19	21	2		
≥45 and <55	63	24	106	26	43	2		
≥55 and <65	93	36	140	35	47	1		
≥65	24	9	40	10	16	1		
							1.21	.89
Gender								
Male	162	63	179	43	17	20		
Female	96	37	231	56	135	14		
Type of employment contract							38.39	.00
Veterinarian-inspector	174	67	268	64	94	3		
Practitioner	84	33	149	35	65	2		
							1.131	.29
Additional employment contract								
Yes	55	32						
No	203	79						
Team								
Living animals	70	27	112	28	42	1		
Abattoirs	152	59	247	59	95	0		
Import	36	14	54	13	18	1		
							.187	.91
Years of employment								
>5	18	7						
≥5 and >10	102	40						
≥10 and >20	54	21						
≥20	84	32						
Proactive behaviour								
Low	43	17						
Medium	169	67						
High	38	15						

3 Based on a list of employees provided by the NVWA in February 2014 which does not include information on years of employment, additional employment contract and proactive behaviour.

4 Veterinarian inspector who coordinates inspecting activities at one or more companies.

5 The function has recently been redefined. This explains why there are more senior veterinary inspectors in the sample than in the population.

6 For example, trainees and employees sent on secondment.

Phrasing of the items: Nine individuals remarked on the formulation of the questionnaire items: some statements were too black-and-white, leaving no opportunity for differentiation, or were formulated rather cryptically. Because of the reactions of some inspectors to the PSM items in the pilot study I assume this critique to be related to the items designed to measure PSM. Also, the PSM items were the only items of the questionnaire in which very general questions were asked. The other questions were much more specific, referring directly to the profession of veterinary inspector. Veterinarians are generally rather down-to-earth and pragmatic and therefore they might have had difficulties identifying with propositions about fuzzy concepts such as 'the common good', 'the good of society' or 'civic duty'. Unfortunately, we cannot trace the comments back to particular items.

Aspects missing from the questionnaire: Sixteen individuals mentioned topics that they would like to have seen in the questionnaire. Some stressed that more attention should have been paid to the welfare of employees. Others remarked that they missed questions about 'communication in daily practice' and 'the general functioning of the NVWA'. These remarks may indicate wrong expectation management on our part. However, it is difficult to trace where these incorrect expectations come from. None of the 'missing' topics were mentioned in the cover letter and neither, as far as I know, did I give anybody cause to assume that the research would touch upon these topics during the interviews.

Appropriateness of the dilemmas: Nine individuals criticized the appropriateness of the cases used in the questionnaire. They pointed out that the dilemmas used in this research focused too narrowly on the work of the 'abattoirs' and 'livestock' teams, whereas no attention was paid to the work of the 'import' team. I think that this is an appropriate remark. However, because the 'import' team was rather small (13 % of entire population), I had decided to leave specific cases characteristic of the work in the 'import' team aside, and instead ask individuals to put themselves in the place of a person faced with the hypothetical dilemmas described.

Length of the questionnaire: Fourteen individuals complained about the length of the questionnaire. Most of them explained that answering all questions took significantly longer than the predicted 20 minutes, a prediction formulated on the basis of the pilot study. Even though I had timed beforehand how long it took to complete the questionnaire, this criticism seems to be justified: on average, respondents spent more than 20 minutes filling in the questionnaire. The exact length of time could not be calculated, because many respondents did not fill in the questionnaire in one go but at different moments. The software we used registered start and finish of the filling-in process without keeping track

of pauses. One reason for this miscalculation might be that the respondents primarily work in the field and have limited experience with computer-related tasks.

Positive feedback: Nineteen individuals were explicitly positive. Some mentioned that they enjoyed filling in the questionnaire and/or that the items gave them food for thought. Others wished us good luck with the study, and almost 50 % (n = 127) left their email addresses because they wanted to receive the findings of the study.

Other remarks: Sixteen respondents had remarks about the questionnaire that could not be fitted into any of the five categories discussed above. Because these remarks were unique and sometimes out of context, I will only mention a few here without discussing them in detail. For example, one respondent said that professional peer-reviews and discussion of critical cases are the most important method to reach uniformity. Another mentioned that practitioners are often used to 'fill the gaps' rather than as fully-fledged employees. A third person pointed out that the goals of the NVWA are clear, but that it is unclear to the employees which methods they are expected to apply.

Table 8 Frequency table of remarks about the questionnaire

	No	**%**
Positive feedback	18	22
Aspects missing from the questionnaire	16	20
Other remarks	16	19
Length of questionnaire	14	17
Appropriateness of the dilemmas	9	11
Phrasing of the items	9	11
Total	82	100

Summing up: Quite a large number of respondents took up our invitation and provided specific remarks about the questionnaire. Most remarks fell into the categories 'positive feedback', 'length of the questionnaire', and 'aspects missing from the questionnaire'. Only nine respondents (3% of the sample) criticized the formulation of some items as either being too 'black-and-white' or too vague. This low percentage may be interpreted as an indication of the high validity of the questionnaire.

6.2.3 Quantitative Analyses

In this subsection I will describe all statistical techniques used to analyse and describe the data, and explain why I used them. First, I discuss the techniques used to assess the construct validity of the measurement instruments. Second, I elaborate on techniques used to explore the data. Finally, I focus on techniques performed to test the hypotheses formulated in the theory chapter. Almost all analyses were performed with SPSS 21 (Statistical Package for Social Science). LISREL 9.1 was only used once: to perform a confirmatory factor analysis. This was necessary because SPSS does not have a function for running a confirmatory factor analysis.

6.2.3.1 Statistical techniques used to assess construct validity

The construct validity of a measurement can be assessed by data-reducing techniques. The instrument measuring PSM has frequently been verified in the past (e.g., Kim et al., 2013; Vandenabeele, 2008) and is highly theory driven. Therefore, I used confirmatory factor analysis to assess the construct validity of the concept of PSM. In contrast, the construct validity of the instrument measuring professional role identity was assessed by a specific exploratory factor analysis – principal component analysis – because the measurement instrument was newly developed for this study and had not yet been tested. Both analyses are discussed in detail.

Confirmatory factor analysis

In order to assess whether the PSM measurement instrument has a good fit with its underlying theoretical model (construct validity), a confirmatory factor analysis (CFA) was performed in Subsection 6.1.2.1. Because the data were normally distributed, the model fit was estimated by the maximum likelihood method. I did not use the frequently cited *chi*-square index as an indicator of good fit, because it has been criticized in structural equation modelling literature (e.g., Brown, 2006; Kelloway, 1989). Following the recommendation of Williams, Vandenberg and Edwards (2009), I used the comparative fit index (CFI) and the root mean square error approximation (RMSEA) as fit indicators. They "are better in assessing model fit" (William et al., 2009, p. 585). According to Williams et al. (2009), a CFI value above .95 and RMSEA indicator below .08 may be interpreted as demonstrating good fit. Using more than one index to examine the overall model fit is recommended, because it is possible for a model to be adequate on one fit index but at the same time inadequate on others (Bollen, 1989).

Principal component analysis

In order to assess the dimensionality and construct validity of the set of items developed to measure professional role identity a principal component analysis (PCA) was performed in Subsection 6.1.2.2. The extraction of principal components was based on the calculation of the Eigenvalues on the basis of the correlation matrix of the dataset. According to Field (2009), even though solutions generated by PCA and exploratory factor analysis – another frequently used data reduction technique – differ little in most circumstances, the choice between these two techniques evokes strong feelings. In this study I decided to use PCA with orthogonal factor rotation because it is a psychometrically sound procedure (Field, 2009). The rotation method applied is Varimax, developed by Kaiser (1958), which is regarded as the best and most widely used rotation method in psychological research (Fabrigar, Wenger, MacCallum & Strahan, 1999). Varimax is based on the idea of maximizing the variance of squared factor loadings across variables.

6.2.3.2 Statistical techniques used to describe the data

I performed a number of analyses in order to get a better understanding of what the data told us. For example, do different groups of respondents have different scores on the core constructs of this study? And how are these constructs interrelated? I here describe the statistical techniques performed to gain answers to these questions.

Pearson's product-moment correlation

The degree of linear relationship between two variables ranking from -1 to +1 (Kutner, Nachtsheim, Netter & William, 2005) is assessed by Pearson's product-moment correlation coefficient. In this study I used it to investigate, for example, how the different dimensions of the concept of professional role identity were correlated, and how they were related to PSM. In Tables 10a and 10b (Chapter 7), a complete overview of the correlations of the core variables under study is provided.

T test

The *t* test is a statistical technique to compare the actual difference between the means of two variables with the variation in the dataset. One of the advantages of the *t* test is its robustness and reliability, even when applied to a relatively small number of cases (Kutner et al., 2005). In this study, the *t* test was used to analyse whether different groups of employees (e.g., men versus women; practitioners versus 'regular' veterinary inspectors) scored differently on the core variables.

Ordinary least square regression analysis

Ordinary least square regression analysis is a common way to predict the values of an outcome variable on the basis of variance in one or more independent variables, assuming that the relationships between the independent and the dependent variables are linear (Kutner et al., 2005). I used this parametric test for example to investigate if there was an effect of number of years of employment on PSM.

6.2.3.3 Statistical technique used to test hypotheses

In order to test hypotheses 1 and 2, *logistic regression analyses* were performed. A logit[7] model is necessary because the independent variable – decision-making – is categorical, which means that the general assumption of linearity in linear regression models between the independent and dependent variables is violated. By expressing the linear regression equation in logarithmic terms, logistic regression analysis avoids the problems associated with violating the assumption of linearity. In logistic regression models a linear relationship between a continuous predictor and the logit of the dependent variable is assumed (Fields, 2009). The relationship between the independent and the outcome variables is expressed as an odds ratio (OR), which is the estimated increase in the logit of the outcomes variable associated with a one-unit increase in the predictor variable.

6.3 Missing data

Systematic lacunae in the dataset are problematic since they may bias the results. For most of the constructs used in this study the percentage of missing values was very small (ranging from 0 to 3.6 %). Only for the dependent variable 'decision-making' were there about 10.5 % missing values. (See Table 9 below for more information.) The relatively high percentage of missing values for this variable is probably related to respondents getting tired of answering questions and therefore quitting before they reached the final questions on decision-making. This assumption is supported by two observations: 1) a number of respondents commented that filling in the questionnaire took too long (see also Section 6.2.2), and 2) the missing values are not random but systematic. Individuals who did not answer Dilemma 1 did not answer Dilemmas 2 and 3 either. Replacing the missing values by the mode makes variance estimates artificially smaller and therefore less valid (Van der Velde et al., 2004). For this reason, all missing values were excluded from the analysis;

7 Alternatively, I could have used a probit model. The choice between logit and probit models is one of convenience (Long, 1997). Since I feel comfortable with the interpretation of logit coefficients as odds ratios, I decided to use a logit model.

hence, correlations were computed only for pairs of non-missing values (pairwise deletion of missing data). For regression analysis the entire case was deleted if any of the variables included in the analysis contained a missing value (list-wise deletion of missing data). The same holds for the factor analyses I performed.

Table 9 Summary of missing values

Concept	No of missing values of N = 258	% of missing values of N = 258
PSM	0	0
Professional role identity		
Commitment to economic interest	0	0
Commitment to animal welfare	9	3.6
Commitment to public health	4	1.6
Strict rule enforcement	3	1.2
Commitment to the inspected	2	.08
Professional identification	0	0
Decision-making Dilemma 1	23	8.9
Decision-making Dilemma 2	27	10.5
Decision-making Dilemma 3	25	9.7

6.4 Limitations of the research design

Some limitations of the research design of this study, which should be considered because they raise a number of issues for future research. The first limitation of this study relates to its generalizability or external validity. The results are based on one group of professionals within one country. This raises the question whether the findings also apply to other countries and professions. In the near future, the results of this study will be compared to a similar survey study conducted in Belgium.

Second, the research design is cross-sectional. All data were collected at one point in time. This provides a threat to internal validity; longitudinal data yield better internal validity because they allow researchers to make stronger causal claims.

Third, the data used in this study were collected from one group of respondents – veterinary inspectors – by means of one questionnaire. Some scholars are sceptical about measuring both the independent and the dependent variable on the basis of just one source (e.g., Podsakoff et al., 2003), since findings might be distorted by *common method bias* (CMB) or *method effects*: the risk that the estimated variance is attributed to the measurement

method rather than to the construct under scrutiny. Other scholars claim that the risk of CMB is overestimated (Lance, Dawson, Birkelbach & Hoffman, 2010). Nevertheless, we put the dependent and independent variables at great distance from each other in the questionnaire. This reduces the risk of respondents cognitively combining the independent and dependent variable, which might trigger a CMB-biased pattern of responses. Ideally, I would have used different sources to measure decision-making (independent variable), such as information offered by colleagues or registered data. Registered data, however, were not made available by the organization, and information provided by colleagues was not a valid source because veterinary inspectors often work on their own; colleagues are therefore not able to judge how a veterinary inspector is likely to act in a dilemma situation.

Another potential risk of self-reported data is that of *social desirability*, which refers to the tendency of respondents to answer all items in a way that presents them in the best possible light (Dooley, 2001). They might do this consciously or unconsciously. Thus, the mean scores presented in this chapter need to be interpreted with some caution. However, since one of the most important aims of this study was to investigate the relationships between two or more variables, and the threat of socially desirable answers is not applicable to this kind of analysis, I consider the risk of such answers to be minor. Nevertheless, this study could have benefitted from including a number of items that control for social desirability. However, since the questionnaire was rather long and the usefulness of social desirability measurements has been doubted (e.g., Borkenau & Ostendorf, 1992), I decided not to include this measurement.

An additional weakness of this study might be that few respondents were asked to fill in the questionnaire who were less familiar with hypothetical dilemma situations included in the survey research: veterinary inspectors of the team 'import'. Respondents of this group might have found it difficult to put themselves in the place of the person being faced with the described dilemma, which could lead to invalid decision-making. The reason why I included this group of participants regardless of this potential weakness was twofold: 1) some respondents used to work in teams in which they used to encounter the dilemmas described; 2) the size of the research population was small. Excluding the respondents of an entire group of respondents would have yielded an even smaller sample size.

Some of the measurements instruments have a Cronbach's α lower than .70, which is usually seen as cut-off score for acceptable reliability (e.g., Hair, Anderson, Tatham & Black, 1998). I decided to hold on to these variables for different reasons. First, the level of Cronbach's α depends on the number of items; the greater the number of items the higher Cronbach's α (e.g., Dooley, 2001; Van der Velde et al., 2004) and all constructs with a low α were measured only by two or three items (all the dimensions of the concept of professional

role identity, the concept of professional identification, and the concept of commitment to the inspectee). Second, the instrument measuring professional identification had been used frequently in previous studies (e.g., Hekman et al., 2009; Loi et al., 2004). Third, the principal component analysis I carried out support the idea that the items measuring professional role identity could be clustered into four different dimensions.

Finally, the dataset on which this study is based is relatively small (N = 258). A larger sample size leads to more accurate parameter estimations and provides more chances to cancel out unrepresentative elements (Dooley, 2001). The entire research population (all veterinary inspectors in the Netherlands), however, consists of only 403 individuals, which makes a large-N study impossible.