Animal Welfare 2008, 17: 289-304
ISSN 0962-7286

RICHPIG: a semantic model to assess enrichment materials for pigs

MBM Bracke

Animal Sciences Group of Wageningen University and Research Centre, PO Box 65, 8200 AB Lelystad, The Netherlands
Email: marc.bracke@wur.nl

Abstract

A computer-based model was constructed to assess enrichment materials (EMats) for intensively-farmed weaned, growing and fattening pigs on a scale from 0 to 10. This model, called RICHPIG, was constructed in order to support the further implementation of EC Directive 2001/93/EC, which states that "pigs must have permanent access to a sufficient quantity of material to enable proper investigation and manipulation activities". This paper describes the underlying conceptual framework for assessing EMats and explains the concepts, procedures and calculation rules used for semantic modelling. A (parsimonious) weighted average calculation rule was used to calculate enrichment scores from assessment criteria scores (which specify welfare relevant material properties of EMats) and weighting factors (WFs, which specify the relative importance of the assessment criteria). In total, 30 assessment criteria were identified and classified as object design criteria (eg novelty and accessibility), behavioural elements (eg nose, root, chew), biological functions (explore and forage), manipulations (ie object-directed behaviours), other (non-manipulative) consequences (eg aggression and stress) and object performance criteria (eg changeability/destructibility and hygiene). WFs were calculated from a systematic analysis of 573 scientific statements collected in the database, using 11 so-called weighting categories (Wcat, ie scientific paradigms to assess welfare such as the study of natural behaviour, consumer demand studies and stress-physiology) to assign Wcat level scores (which indicate the intensity, duration and incidence of a welfare impact) to the assessment criteria. The main advantages of the RICHPIG model are that it is based explicitly on available scientific information, that it has an explicitly formulated conceptual framework, is transparent, disputable, upgradeable, robust and reasonably in accordance with expert opinion. Major scope for improvements exist in the form of the need for further upgrading with new knowledge, empirical validation and (further) implementation in political decision-making processes.

Keywords: *animal welfare, behaviour, decision support system, environmental enrichment, pigs, semantic modelling*

Introduction

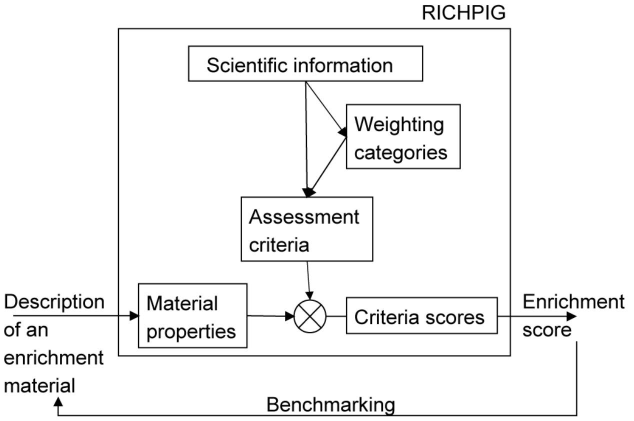
The growing field of applied ethology, which deals with assessing farm animal welfare, does not always provide the required answers for science-based political decision making. 'Semantic modelling' tries to help bridge the gap by developing procedures and computer-based tools (so-called decision support systems) for quantifying welfare assessment based on the available scientific literature. In brief, semantic modelling can be defined as formalised welfare assessment based on available scientific information.

The methodology was originally developed for overall welfare assessment at system level in dry sows, showing that semantic modelling was possible in principle and that it correlated reasonably well with expert opinion (SOWEL), (Bracke 2001; Bracke *et al* 2002a, b). Subsequent work on tail biting in pigs showed that the method could also be applied to other, related issues and that validation could be performed empirically (PIGTAIL), (Bracke *et al* 2004a, b). An independent research group constructed a semantic model to evaluate newly-designed housing systems for laying hens (FOWEL), (De Mol *et al* 2006). The most recent model was designed to assess enRICHment materials for PIGs. It is called RICHPIG. This model has a refined underlying conceptual framework, which resulted in a new classification of criteria to assess enrichment value and it has improved calculation rules, eg taking into account the levels of (un)certainty of the underlying scientific information.

RICHPIG was constructed to support the Dutch Ministry of Agriculture with the further implementation of EC Directive 2001/93/EC. This directive states that: "Pigs must have permanent access to a sufficient quantity of material to enable proper investigation and manipulation activities, such as straw, hay, wood, sawdust, mushroom compost, peat or a mixture of such, which does not compromise the health of the animals". The interpretation of this directive was not straightforward due to words like 'such as', 'proper' and 'sufficient'. The modelling was intended, more specifically, to generate operational assessment criteria that determine the enrichment value of toys and substrates for intensively-housed pigs.

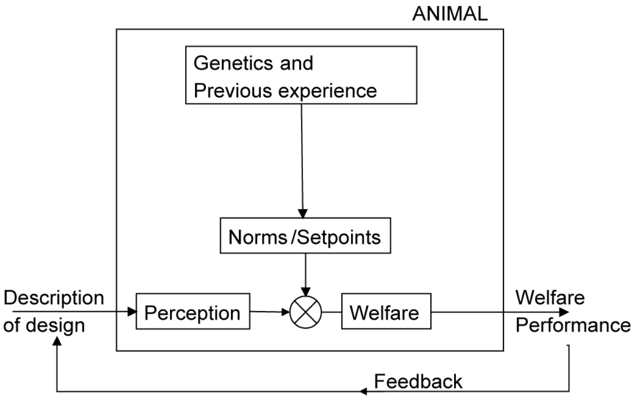
https://doi.org/10.1017/S0962728600032218 Published online by Cambridge University Press

Figure 1



Overview of RICHPIG's structure showing the assessment of a new enrichment material on the horizontal axis and the modelling procedure on the vertical axis.

Figure 2



Conceptual framework for welfare assessment (adapted from Wiepkema 1987 and Anonymous 2001a).

Spoolder *et al* (2003) formulated criteria to evaluate methods for integrated welfare assessment; a model should be valid, reliable and feasible. Validity implies that the model should be in accordance with the present state in (biological) science. To this end, the model should be objective, reflecting what (is known about what) matters to the animals, rather than reflecting the personal opinion of the modeller, and it should be transparent, such that it can be disputed and, if necessary, upgraded.

The objective of this paper was to explain RICHPIG's structure in the hope of allowing critical evaluation, identification of the scope for further improvements and implementation in (policy-making) practice.

Overview of RICHPIG

The model was designed to formally derive enrichment scores for enrichment materials (EMats) on a scale from 0 (worst) to 10 (best) from scientific information collected in a database.

Figure 1 shows RICHPIG's main components and procedures. RICHPIG is a decision support system programmed as a so-called relational database in the software Microsoft Access and Excel.

In order to assess a new EMat the user must assign a set of criteria scores (one for each assessment criterion) based on the material properties of the EMat. RICHPIG then multiplies the criteria scores and pre-determined weighting factors (WF) to generate an enrichment score. The user is assisted in assigning criteria scores and in interpreting the enrichment score by a 'feedback' loop called 'benchmarking', which involves a comparison with benchmark (example) EMats in the model.

For each assessment criterion a WF was calculated based on a formal analysis of the scientific information using so-called weighting categories (Wcat) such as (the study of) natural behaviour, preferences, frustration/aversion, production, health and fitness.

RICHPIG's structure (Figure 1) closely resembles its underlying biological conceptual framework for animal welfare assessment (Figure 2). This follows from the common objective of assessing welfare as what matters to the animals from their point of view (Bracke *et al* 1999a, b, c). The conceptual framework and RICHPIG's components (EMat treatments, calculation rules, assessment criteria, weighting factors, weighting categories, scientific statements and scientific references) will be described in more detail below and include a reference to the underlying principles, main shortcomings and scope for improvement.

EMat treatments used for benchmarking

The model was designed for intensively-housed, weaned, growing and fattening pigs that are kept in a so-called 'reference pen' without EMat (Bracke *et al* 2006; Table 1). The provision of an EMat in the reference pen was called an 'EMat treatment'. EMat treatments were included in RICHPIG for benchmarking, ie to 'see' the model work and to provide operational support for the assessment of new EMats.

In the model 130 different EMat treatments were listed. They were derived from the literature and from the need to illustrate assessment principles. A large subset of 74 EMats were derived from van de Weerd *et al* (2003b) and used for the 'empirical' validation (described below). Another (slightly overlapping) subset of 64 EMats were used for validation against expert opinion (Bracke *et al* 2007a).

The EMat treatments were described with reference to the type of material, its dimensions, amount, provision frequency, position in the pen and the pigs' age/size for which the EMat was designed. Examples of EMat treatment descriptions and calculated enrichment scores are given in Table 1.

https://doi.org/10.1017/S0962728600032218 Published online by Cambridge University Press

Table 1 Example descriptions and enrichment scores of several enrichment materials in RICHPIG.

Enrichment material	Score	Description
Reference pen (no enrichment)	1.46	Pen without enrichment material, otherwise (just) meeting minimum legal requirements for animal welfare. Typical/standard pen for weaners (as of 25 kg), growers and fatteners (up to 100 kg), respectively. Pen surface for fatteners 0.7–1 m^{-2} pig. Pigs typically fed *ad libidum* pellets, partly slatted concrete floor, stable group of approximately 10 pigs per pen.
Metal chain	2.24	A metal chain, hung vertically, at shoulder height, some 20 cm off the back of the pen.
Plastic ball	2.32	Heavy plastic ball (35 cm diameter) free on the pen floor.
Rubber hose cross	3.04	Two rubber hoses, fixed in the form of a cross, suspended on a chain, slightly above shoulder height.
Rope	3.29	Straight sash cord (cotton 1 cm diameter, 40 cm long) suspended from the pen gate at shoulder height (daily) adjusted according to consumption.
Pinewood beam	4.25	Pinewood beam (13 cm diameter, 1.5 m long) suspended by chains to the wall, at 'knee' (carpus) height.
Earth	4.71	Earth in a small trough (dimensions: 15 × 20 cm).
Foodball	5.20	The Edinburgh Foodball® , containing food pellets that drop out when the ball is rooted upon (refilled once daily).
Mushroom compost	6.53	Spent mushroom compost on a horizontal metal rack (1 m^2 above the pigs' heads), grid size 30 mm^2, compost refreshed daily, approximately 1/3 kg pig^{-1} day^{-1}.
Strawrack device	6.54	Coarse chopped straw from a rack with a trough, a chain (to facilitate sliding of the straw) and a soft-wood beam (8 cm diameter, 50 cm long) hung horizontally above the trough on two chains (straw use: 10–20 g pig^{-1} day^{-1}; straw length: 11 cm).
Straw twice daily	7.08	A handful of long straw provided twice daily (approximately 20 g pig^{-1} day^{-1}).
Fodderbeets	7.09	Roughage, chopped fodderbeets (low DM) in a trough, provided *ad libidum* once daily.
Long straw and branches	8.34	Long straw provided once daily in a pen with two fir branches (which are renewed every month or when destroyed).
Straw and beet roots	8.54	When whole straw mixed with chopped beet roots provided *ad libidum* on the pen floor once daily.

An enrichment score expresses an EMat's enrichment value on a scale from 0 (worst) to 10 (best). This formally-derived score is designed to express the overall level of satisfaction and/or frustration of the pigs' (welfare) needs to explore, play and forage.

When we compared the preliminary model's overall enrichment scores for 64 EMats with expert opinion (senior applied ethologists presented with the EMat descriptions and their model scores) a high correlation ($r = 0.97$, $P < 0.05$) was found (Bracke *et al* 2007a). The present model had a slightly lower, but still very high correlation ($r = 0.96$, $P < 0.05$).

How enrichment scores were calculated

Enrichment scores were calculated from so-called assessment criteria scores and their weighting factors (WFs) using a weighted average calculation rule. This rule had been used successfully earlier (Bracke *et al* 2002a); it is very simple (ie parsimonious) and it has been widely used (eg in education).

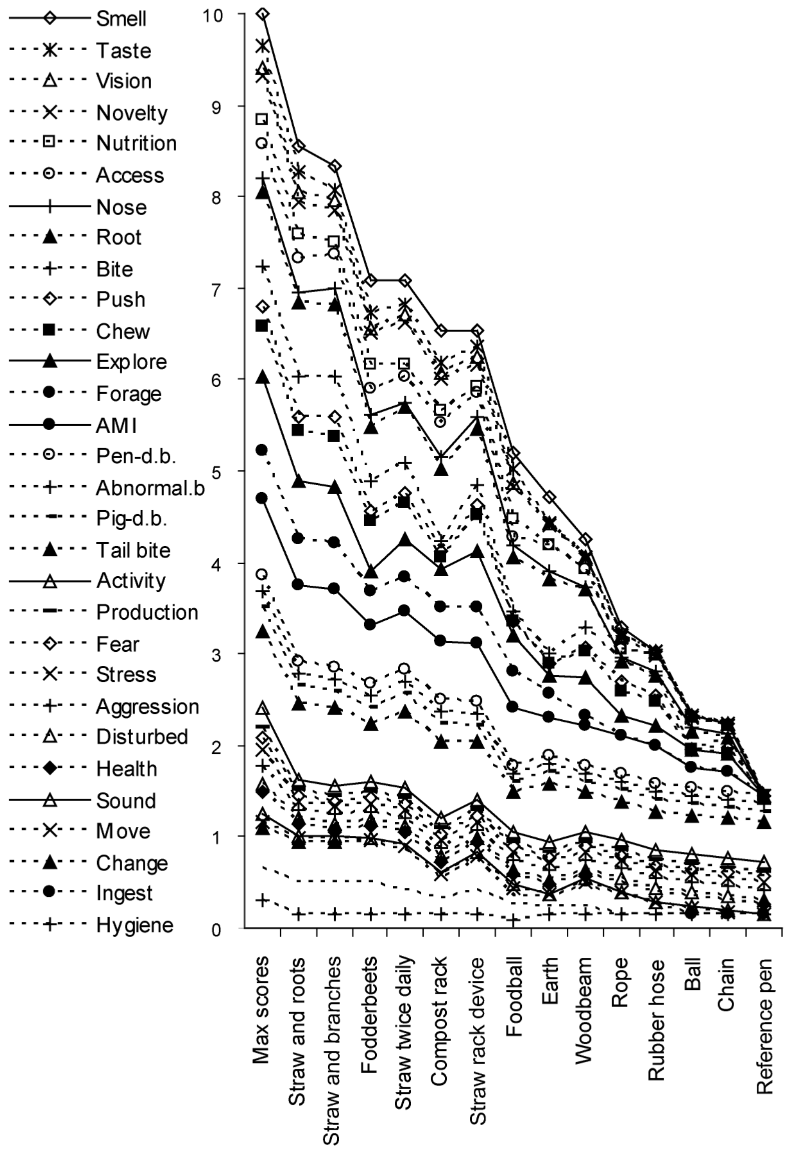
Figure 3 shows the breakdown of the enrichment scores of the example EMats introduced in Table 1. Each assessment criterion contributes to the overall scores (upper line) with normalised assessment criteria scorings which are the products of the normalised WFs and the assessment criteria scores. The assessment criteria scores (scale 0 to 10, see Table 2) represent the degree to which the EMat fulfils the criteria. The normalised WFs represent the proportional importance of the assessment criterion relative to the other assessment criteria in the model. Normalised WFs for all assessment criteria together add up to 1 and are shown in Figure 3 under the label 'Max scores', which add up to a total enrichment score of 10 and where, by definition, an assessment criteria score of 10 was given for each assessment criterion. Normalised WFs are the products of the WF of an assessment criterion divided by the sum of WFs of all assessment criteria in the model.

Since the normalised assessment criteria scorings in Figure 3 (which range between 0 and normalised WF for each criterion) have been stacked, the lines can converge, but they cannot cross each other. This follows from the weighted average calculation rule to calculate overall enrichment values.

The relatively large distances between one assessment criterion's line and the previous, lower line in Figure 3 shows, for example, that 'earth' benefited considerably from being rootable, compared to its adjacent EMat treatments ('foodball' and 'woodbeam'). In accordance with current expert opinion (Bracke 2006), it also shows that the

https://doi.org/10.1017/S0962728600032218 Published online by Cambridge University Press

Figure 3



Normalised assessment criteria scorings for the example enrichment material (EMat) treatments described in Table 1 and stacked in the order presented in the legend. Max scores: Ideal EMat treatment with a total enrichment score of 10 and maximum criteria scores of 10 for each assessment criterion. Solid symbols indicate the more important assessment criteria which, by definition, have a relatively large distance to the previous, lower criterion in the legend, at least for 'Max scores'. Classes of assessment criteria (see text below) have been indicated with a solid line for the first item of each group. d: 'directed'; b: behaviour; AMI: animal-material interaction.

better, straw-based EMats, in general, benefited from being ingestible, changeable (ie destructible), chewable, interactible (AMI), explorable, forageable and rootable (see later discussion of the choice of the assessment criteria).

Figure 3 illustrates, thus, that overall enrichment scores for EMats derive from the degree to which each criterion is fulfilled (criteria scores) and from the relative importance of each criterion (normalised WFs).

Table 2 shows some (unweighted) assessment criteria scores for the example EMats. They were derived from an evaluation of the material properties of the EMats in relation to the science-based assessment criteria (see below).

These scores were assigned on a scale from 0 to 10, preferably using the 5 points 0, 2.5, 5, 7.5 and 10, indicating minimal (0) to maximal (10) satisfaction of the criterion. The scores in Table 2 have been used in Figure 3. For example, the first score of 7.5 shown in Table 2 (for EMat 'straw and roots' and 'smell', which expresses the degree to which the EMat has an attractive smell for the pigs) is represented in the upper left corner of Figure 3 as 0.75 times the distance for 'Max scores' between the lines for 'smell' and 'taste'. A more detailed explanation of the assessment criteria (scores) however, requires, first, a more detailed description of the underlying conceptual framework.

Conceptual framework for welfare assessment

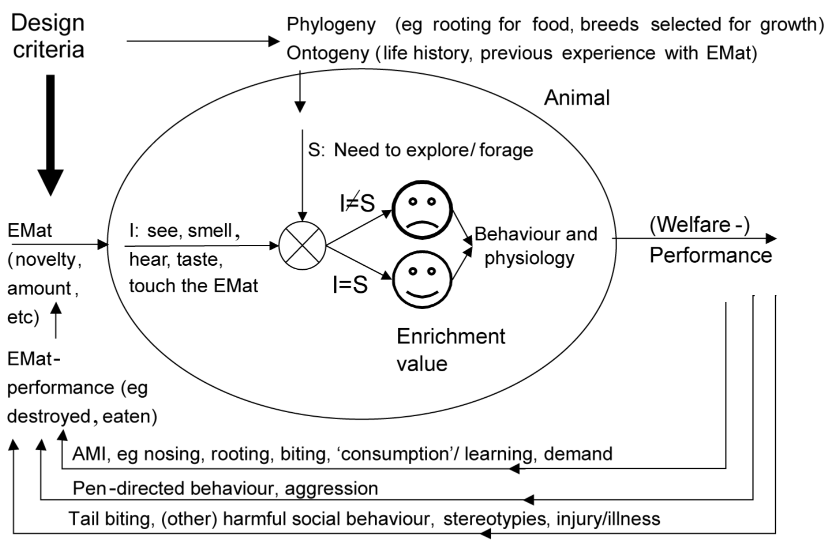
The conceptual framework for modelling environmental enrichment in pigs derived from Wiepkema (1987). It was previously formulated as a commonly-used framework for animal welfare assessment in Anonymous (2001a), had been used for semantic modelling of welfare in pregnant

Table 2 Examples of assessment criteria and their scores for enrichment materials (EMats) in RICHPIG.

EMats	Smell	Novelty	Access	Root	Chew	Change	Move
Straw and roots	7.5	7.5	10	10	10	10	7.5
Straw and branches	7.5	7.5	10	10	10	10	7.5
Fodderbeets	10	7.5	7.5	7.5	10	10	5
Straw twice daily	7.5	10	7.5	7.5	7.5	10	7.5
Compost rack	10	7.5	10	10	2.5	5	5
Strawrack device	5	5	7.5	7.5	7.5	7.5	5
Foodball	5	7.5	2.5	7.5	2.5	2.5	10
Earth	7.5	5	7.5	10	2.5	2.5	2.5
Woodbeam	5	2.5	5	5	5	5	5
Rope	2.5	2.5	2.5	0	5	5	2.5
Rubber hose	0	0	5	0	5	2.5	2.5
Ball	0	0	2.5	2.5	0	0	7.5
Chain	0	0	2.5	0	0	0	2.5
Reference pen	0	0	0	0	0	0	0

Figure 4

Schematic representation of the conceptual framework for assessing environmental enrichment for pigs. EMat: enrichment material; AMI: animal-material interactions; I: Istwert, the environment as perceived by the animal; S: Sollwert, setpoint or norm of the animal.



sows (Bracke *et al* 2002a, b) and was described in relation to natural behaviour in Bracke and Hopster (2006).

Figure 4 shows the basis for the conceptualisation of the (classes of) assessment criteria in the model. 'Design criteria' represent the environment in the widest possible sense. It includes both the here-and-now environment as it impinges upon the animal (Istwert, reference pen with/without EMat, conspecifics, stockpersons) and the past environments (including both the animal's ontogeny/life history and its phylogeny/evolutionary history). The past environments co-determine the norms/Sollwerte of the animal here and now (through, eg natural selection and conditioning). The here-and-now environment is perceived by the animal through its perceptual system, which includes its ability to smell, see, hear, taste and touch.

The animal's behavioural and physiological responses form its (welfare) performance output which can be measured and grouped into a number of motivational systems/welfare needs (Bracke *et al* 1999c) that involve emotions and that are regulated with feedback loops (making predictability and controllability so important for animal welfare), (Wiepkema & Koolhaas 1993).

Within the conceptual framework, the concept of 'proper investigation and manipulation' as used in directive 2001/93/EC, implies functional behaviours that meet the animal's 'enrichment' needs, especially its ethological needs to explore, play and forage, and its associated emotions of curiosity and boredom, while other needs were excluded from being modelled. Pigs are motivated to explore, play and forage and they need objects or substrates to satisfy these needs (eg Wood-Gush & Vestergaard 1991). More satisfaction may be anticipated when behaviours have functional consequences, eg when they are performed as a functional chain of behaviours (eg rooting is more valuable when it leads to obtaining food). In general, the more the animals are motivated to perform the (various component) behaviours and the more severe the negative consequences when prevented performing the behaviours (eg increased tail and ear biting), the more important these behaviours must be to the animals. Accordingly, weighting factors (WF) can be derived from the intensity, duration and incidence of the behavioural and physiological (including stress, reproductive and patho-physiological) responses and their biological significance.

The feedback loops in Figure 4 show different stages of the animal's engagement with EMats. For example, when an animal is provided with an EMat (design criterion) the EMat will be explored (performance, AMI). When EMats are absent or inadequate, the animal may show primary feedback, eg search behaviour and pen-directed behaviour. When it fails to find a suitable substrate, secondary strategies will develop such as tail biting and/or stereotypic oral behaviour patterns. Finally, the problem may escalate, eg when cannibalism develops.

The conceptual framework thus identifies different sources of information that are relevant to assess environmental enrichment for pigs. Starting from the framework's input side these include object design criteria, behavioural elements, biological functions, manipulations (of EMats, pen and animals), other (non-manipulative) consequences and object performance criteria. These define the 6 classes of the assessment criteria in the model.

Assessment criteria

'Assessment criteria' are material properties, ie descriptors of an EMat, that may be regarded as factors contributing (additively) to welfare/enrichment value, ie 'proper investigation and manipulation'. Assessment criteria include both design criteria and performance criteria. Note that these are related by the feedback in the conceptual framework and this allows, for example, the translation of the statement 'straw elicits rooting', where 'rooting' is a behavioural, animal-based performance criterion, into the statement 'straw is a rootable material', where 'rootability' is an environment-based design property of the material 'straw'. In total 30 assessment criteria were formulated in RICHPIG.

The class of object design criteria includes 6 different perceptual causes of enrichment: 'smell', 'taste', 'visibility', 'novelty', 'nutritious' and 'accessible'. These items were formulated based on an analysis of the pigs' senses ('smell', 'taste', 'see', 'hear' and 'touch'), an attempt to avoid overlap with other criteria (eg 'touch' was already covered by the ethogram-based criteria 'nose', 'bite' and 'chew'; and 'hearing' was subsumed under 'sound producing', which was classified as object performance criterion, see below) and aspects encountered in the literature specifically recognised as being important (eg 'novelty', 'being nutritious' and 'accessible').

Behavioural elements form a class of 5 assessment criteria that are ethological elements of 'proper investigation and manipulation'. These are: 'nose', 'root', 'bite', 'pull' and 'chew'. This included the main behavioural elements. It excluded minor elements such as 'paw' as these lacked scientific support in the RICHPIG database, eg indicating lack of evidence that 'pawability' significantly contributed to overall enrichment value.

Biological functions form a class of 2 assessment criteria identifying enrichment goals (Sollwerte). These are: 'explore/learn' and 'forage'. Here 'play', which was restricted to 'object play' (as a result of the definition of the domain of investigation in relation to EMats), was subsumed under 'explore/learn'.

Manipulations form a class of 5 assessment criteria that refer to interactions of the animals with the EMat, the pen and penmates in relation to (seeking) enrichment. These are: 'animal-material interactions' (AMI), 'pen-directed behaviour', 'abnormal behaviour' (eg belly nosing), 'tail (and ear) biting' and (other) 'penmate-directed behaviour'.

Other (non-manipulative) consequences form a class of 7 assessment criteria that are non-manipulative (potential) consequences of (the lack of) enrichment. These are: 'activity', 'production', 'fear', 'stress', 'aggression', 'disturbance' (as when pigs playing with a chain on the floor disturb the resting behaviour of penmates) and 'health'.

Object performance criteria form a class of 5 assessment criteria that refer to the way the EMat responds to manipulation. Such responses are indicative of a certain level of control by the pig. Object performance criteria are: 'sound', 'move', 'change' (which includes flexibility and destructibility of the EMat), 'ingest' and 'hygiene'.

Each assessment criterion specifies some aspect that contributes additively to enrichment value. This implies that, contrary to what the term 'criterion' may suggest, a single assessment criterion as used here, is neither necessary nor sufficient for overall enrichment. This is because in the model, enrichment scores are calculated as a (weighted average) function of the degrees of satisfaction of all assessment criteria together. Each assessment criterion is only necessary and sufficient to determine the component of enrichment value it represents.

The formulation of the assessment criteria started with an inventory of the measures used for studying EMats for pigs (Bracke *et al* 2006). In an email questionnaire, experts were also asked to identify material properties/criteria they considered important to assess the enrichment value for pigs (Bracke 2006). Subsequently, the lists were related to the conceptual framework for welfare assessment. Finally, the assessment criteria were shaped further during modelling

https://doi.org/10.1017/S0962728600032218 Published online by Cambridge University Press

when they were applied to many (130) EMat treatments and when they were functionally related to the scientific statements in the database, following a procedure derived from Bracke *et al* (2002a); (where an 'assessment criterion' was called an 'attribute') and described in more detail below.

Together, the assessment criteria in the model were intended to cover all aspects that contribute to enrichment, according to the conceptual framework and according to the scientific information collected in the database, and they were formulated in such a way that their scores could be 'added up' to determine the overall enrichment value/score of the EMat.

Formally, each assessment criterion contains two or more levels that can be ranked from worst to best and that together match the range of conditions (to be) encountered in the assessment domain (here, the range of EMat treatments for which the model is designed). Each criterion, therefore, can be applied to every EMat in the assessment domain, by declaring identity to exactly one level for each criterion (ie criteria apply across EMats). This makes enrichment calculations possible since each ranked level has an assessment criterion (level) score attached to it (eg 0, 2.5, 5, 7.5 and 10 for a criterion with five levels ranked from worst to best), expressing the within-criterion rank. For example, the criterion 'rootability' was assigned the level 'not rootable' (score 0) for the treatment 'rubber mat' and 'highly rootable' (score 10) for the treatment 'earth trough' (see Table 2). Some criteria scales were also reversed. For example, the 'tail biting' scale was reversed because higher levels of tail biting indicate lower levels of welfare/enrichment.

The procedure ensured that the assessment criteria scores were normalised 'from the animal's point of view'. For example, the criterion 'change' (destructibility) does not imply that the more destructible the better for enrichment (as materials can also be too destructible for the pig).

Whenever possible, various related aspects of an assessment criterion were grouped into one concept. For example, measures of the duration and frequencies of tail-biting behaviour as well as tail lesion scores, all loaded onto the criterion 'tail biting'. Not all 30 assessment criteria formulated in RICHPIG were functional for calculating enrichment scores. When adequate information was lacking to assign scores to the benchmark EMats for an assessment criterion, then default values were given, sometimes even for all EMats. A default assessment criteria score of '5' was given when it was expected that future information could be associated with both higher and lower levels compared to the level in the reference pen. The development of animal-based measures for on-farm welfare assessment was not an objective of RICHPIG's construction but would certainly be a welcome supplement of the model as well as providing a potential instrument for empirical model validation. The use of default values resulted in parallel lines shown in Figure 3 (and as detailed below).

Functional sets of assessment criteria scores were given for the classes 'object 'design' criteria (perceptions)', 'behavioural elements', 'biological functions' and 'object performance criteria'. Out of the class 'manipulations' only the assessment criterion 'AMI' was fully used. The other criteria in this class and the class 'other (non-manipulative) consequences' was used only sparsely (as little was known), while the 5 criteria 'activity', 'production', 'fear', 'stress' and 'aggression' were not used, receiving a default score of '5' for all EMat treatments.

A factor analysis of the assessment criteria scores of all EMats in the model (PCA, using SPSS 13.0) (Anonymous 2001b) revealed that 4 factors explained 56.6, 8.7, 6.4 and 5.0% of the variation.

Factor 1, which could perhaps be labelled 'manipulation', contained all 6 perceptual causes of enrichment (with coefficients between 0.59 and 0.82), two behavioural elements ('nose' and 'root', with coefficients of 0.70 and 0.87, respectively), each of the two biological functions ('explore/learn' and 'forage', with coefficients of 0.58 and 0.86, respectively), all 5 criteria in the class 'manipulations' (with coefficients between 0.69 and 0.90) and it contained the criteria 'disturbance' (0.81), 'change' (0.53) and 'ingestion' (0.73).

Factor 2, which could perhaps be labelled 'biteability', contained the criteria 'bite', 'pull', 'chew' and 'change' (with coefficients between 0.73 and 0.86) and, to a lesser extent, 'AMI' and 'ingest' (with coefficients of 0.53 and 0.56, respectively).

Factor 3 contained the criteria 'sound' (0.71), 'move' (0.73) and the negatively loading 'hygiene' (–0.68). This factor could be called 'loose', where the negative loading of 'hygiene' can be explained as the result of reduced 'hygiene', ie enhanced soiling with excreta, due to being 'loose' in the pen.

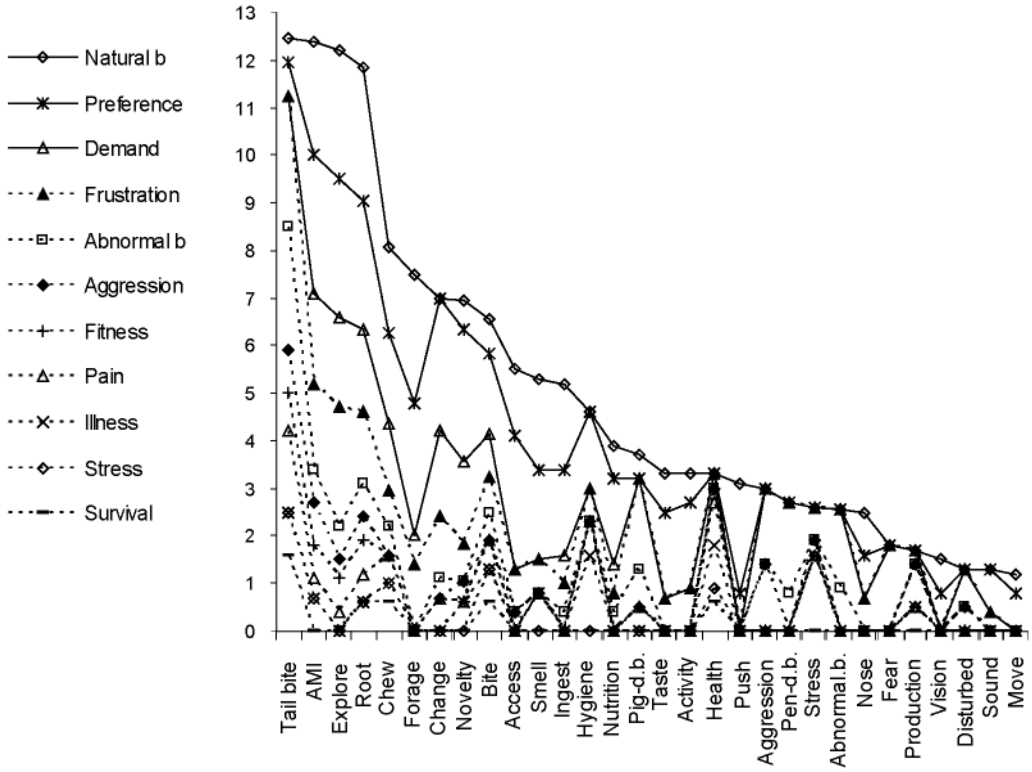
Factor 4 contained the criterion 'health' (0.89).

A multiple-stepwise regression analysis showed that 'forage', 'change', 'AMI', 'root', 'nutritious', 'pull', 'tail bite', 'chew', 'explore/learn', 'novelty', 'bite', 'ingest' and 'accessibility' were subsequently selected.

While the factor analysis identified sets of criteria that point in a similar direction, the stepwise regression identified the criteria that most contributed to the overall scores. Both analyses provide scope for further simplification of the model. At present, however, no attempts were made to further reduce the model because it was considered most important to maintain the full relationship with the underlying conceptual framework and because application of the model has already been made fairly simple.

When assessing a new EMat using RICHPIG the user must specify its relevant material properties. This requires assigning one assessment criteria score (scale 0 to 10) to each of the 30 assessment criteria in the model (including a default score of 5 for each of the 5 non-functional criteria). This task is not as difficult as it may seem, as the user must select and copy the scores of the most similar benchmark EMats already scored in the model. This is because, for optimal scoring and maintaining the model's integrity/coherence, the user must treat similar cases

Figure 5



Absolute scores of (positive and negative) maxima of weighting category level scores for each assessment criterion in the RICHPIG model. The criteria have been sorted according to their overall weighting factors (WF) from left to right. The weighting categories (Wcat) have been stacked in the order of presentation in the legend. The 3 upper Wcat all load positively and are identified by solid lines. The remaining Wcat load negatively and are identified by dotted lines. 'd': directed; 'b': behaviour.

similarly, ie he/she must follow the scorings of similar cases, unless there are good reasons to divert from that scoring. This reduces both the knowledge requirements demanded from the user as well as the potential impact of his/her subjective opinions and interests.

Weighting factors were derived from weighting categories

In RICHPIG the relative importance of each criterion for assessing enrichment value (compared to the other criteria in the model) was expressed with a weighting factor (WF).

WFs were derived from so-called weighting categories (Wcats). The 11 Wcats formulated in RICHPIG were all (higher-order classes of) performance criteria that refer to the different methodologies/paradigms used for studying animal welfare, including the study of natural behaviour, preference tests, consumer demand studies, the study of abnormal behaviour, stress physiology, etc. The Wcats derived from previous work, where they were functional for deriving WFs for assessing the overall welfare status of pregnant sows (Bracke *et al* 2002a, b). In RICHPIG each Wcat had 3 levels (ranked and scored from 1 to 3) which were attributed to assessment criteria in the analysis of the scientific statements in the database. These Wcat level scores were assigned on the basis of an interpretation of scientific statements reporting a given level (in terms of intensity, duration and incidence) of some measure taken within a Wcat study paradigm.

WFs were multiplied with assessment criteria scores to calculate overall enrichment scores. WFs were themselves calculated from so-called weighting category level scores that were assigned when the scientific information on consequences for different aspects of enrichment value (pig welfare) in the database was systematically analysed during modelling (as will be explained in the sections below).

Figure 5 shows the WFs for the assessment criteria in the model, as well as how the scores were built up from component scores for underlying weighting categories (Wcat). For example, 'tail bite' had the highest WF (12.5), which is represented in the figure as the stacking of the absolute values of the negative scores (–11.3) and the positive scores (1.2). The figure shows, for example, that the criterion 'tail bite' generates a high WF mainly due to negative loadings (dotted lines), while the next important criterion 'AMI' generates its WF mainly due to positive loadings.

The resemblance between Figures 3 and 5 is not a coincidence. The enrichment scores in Figure 3 are derived from

https://doi.org/10.1017/S0962728600032218 Published online by Cambridge University Press

stacking weighted assessment criteria scores, while the WFs in Figure 5 are derived from stacking (maxima of) Wcat level scores. In fact, Wcats and assessment criteria serve a similar function at a different level of abstraction. The WFs in both figures are mathematically related, in that the normalised WF for (the 'Max scores' of) 'tail bite', for example, in Figure 3 is the total score shown in Figure 5 (ie 12.5) divided by the sum of the WFs of all assessment criteria (ie 12.5 for 'tail bite' + 12.4 for AMI + 12.2 for 'explore/learn' + … + 1.2 for 'move').

Figure 5 does not show the 'Max scores', as the Wcat were all scored on a 3-point scale ('low', 'moderate' and 'high'), ie +3 for each positive Wcat (ie 'natural behaviour', 'preference' and 'demand') and –3 for each negative Wcat (ie 'frustration/aversion', 'abnormal behaviour', 'aggression', 'reduced fitness', 'pain', 'illness', 'stress' and 'reduced survival'). The absolute total of 33 (11 Wcat × 3 points) is much higher than the highest WF shown in Figure 5 (12.5 for 'tail bite') because the Wcat level scores (on the 3-point scales) were corrected for uncertainty and because the highest possible score (3 or –3) was not often supported by the underlying scientific information (see below).

The WFs were found to be moderately correlated to expert opinion (Spearman Rank = 0.63, $P < 0.001$; Bracke *et al* 2007b).

A more detailed explanation of obtaining the (positive and negative) loadings will be given below, after an introduction of the scientific information in the database.

Scientific references

Scientific references were regarded as the primary source of information for modelling because they provided the most reliable and valid information available.

Scientific references were classified broadly as methodological publications and publications on environmental enrichment. Methodological publications included papers describing principles of semantic modelling (eg Bracke *et al* 2002a, 2004a) and papers describing the biological conceptual framework that made an interpretation of the (welfare-relevant) 'facts' possible (eg Anonymous 2001a).

Publications on environmental enrichment contained 'facts' in the form of scientific statements that were relevant for assessing EMats for pigs. Publications on enrichment were grouped into 3 so-called 'paper classes': A) Full scientific papers and dissertations reporting unconfounded studies providing EMats for intensively-farmed weaned, growing, or fattening pigs. Class A references were analysed in detail, extracting all reported and relevant scientific results and scientific statements; B) Abstracts or proceedings papers reporting unconfounded effects of EMats for intensively farmed pigs. Class B references were analysed for reported results and major statements and C) Other review-type papers and conceptual papers with high standing in science (eg Fraser [1987], on the taste of blood), papers on organic pigs (eg Olsen *et al* 2000) and relevant reviews (eg Sambrook & Buchanan-Smith 1997; SVC 1997). Class C references were analysed for major statements.

Scientific publications were searched using WebSpirs 5.0 with the Cab-Abstract databases from January 1972 to December 2004, using keywords such as 'enrichment', 'toys' and 'pigs', supplemented with searches on the website of the British Society of Animal Science (BSAS) and an email questionnaire asking experts for additional publications (Bracke 2006).

Retrieved papers were analysed, starting with the identification of (potentially) relevant scientific statements (as mainly found in the 'Introduction' and 'Discussion' sections of the papers), scientific results (as described in the 'Materials and methods', and in the 'Results' sections) and new scientific papers (in the 'References' list).

Although paper selection stopped in December 2004, progressive analysis of retrieved papers resulted in a certain degree of saturation of information, whereby new statements often appeared repetitious. This implied that the latter statements appeared to be of limited value in upgrading the (formulations of the) assessment criteria and (the calculations of the) weighting factors (WFs).

The total number of potentially-relevant papers was considerable. More than 100 references were registered. Since full analysis of all references was not feasible, within the constraints of the project, only 55 references were analysed in detail. The remaining papers could, perhaps, be used for analysis at a later date (eg for validation; testing the robustness of the model in the light of this 'new' information).

From the 55 analysed references, 7 were later discarded for formal reasons (paper classification), leaving 48 references generating scientific statements that were transcribed into RICHPIG.

Class A references generating scientific statements were the following (where the number between brackets represents the number of scientific statements incorporated into RICHPIG's database): Apple & Craig 1992 (8); Beattie *et al* 2001 (20); Blackshaw *et al* 1997 (12); Bolhuis *et al* 2003 (4); Courboulay *et al* 2004 (9); Day *et al* 2002a (22); Day *et al* 2002b (8); Feddes & Fraser 1994 (21); Fraser *et al* 1991 (8); Hill *et al* 1998 (14); Jorgensen 2003 (7); Krötzl *et al* 1993 (15); Lyons *et al* 1995 (20); McKinnon *et al* 1989 (11); Pearce & Paterson 1993 (17); Pearce *et al* 1989 (20); Petersen *et al* 1995 (20); Sambraus & Kuchenhoff 1992 (26); Schaefer *et al* 1990 (9); Stubbe 2000 (107); van de Weerd *et al* 2003b (62); van Putten 1980 (18); van Rooijen 1981 (5); Wood-Gush & Beilharz 1983 (13).

Class B references were: Arey & Maw 1995 (4); Beattie *et al* 1996 (13); Bøe 1992 (4); Buré *et al* 1983 (16); Day *et al* 2001 (28); Gonyou & Bench 2002 (1); Grandin & Curtis 1984 (2); Grandin *et al* 1987 (1); Heizmann *et al* 1988 (15); Höges 1991 (6); Horrell & Ness 1995 (1); Kress *et al* 1999 (2); Meyer *et al* 1984 (1); Moore *et al* 1994 (1); van de Weerd *et al* 2003a (3); van Rooijen 1982 (3); Zonderland *et al* 2003a (2); Zonderland *et al* 2003b (4).

Class C references were: Beattie *et al* 1998 (2); Fraser 1987 (26); Olsen *et al* 2000 (19); Olsen *et al* 2002 (2); Sambrook & Buchanan-Smith 1997 (2); SVC 1997 (45).

Scientific statements

A scientific statement describes a scientific finding, observation and/or explanation concerning environmental enrichment in pigs. Scientific statements typically specify relationships between design criteria and (welfare) performance criteria, eg 'Straw provided twice daily decreased tail biting compared to a straw rack, rubber hose or metal chain' (Zonderland *et al* 2003a). The statements were transcribed as much as possible from the formulation used in the source document, supplemented with context-relevant information required for 'stand-alone' intellegibility. Scientific statements were subsequently analysed for information they contained on assessment criteria, Wcats and EMats. Pure opinions from authors, eg about what constituted suitable EMats, were not analysed.

Two randomly chosen statements may serve as an illustration: Hill *et al* (1998) found that growers/finishers manipulated a rubber hose considerably more than a chain, whereas in weaners there was no difference. This was presumed to be due to the relatively large diameter hose (Hill *et al* 1998, p 66); Beattie *et al* (1993) showed that growers housed indoors and given a rooting substrate, such as peat, increased exploration and decreased inactivity and penmate-directed behaviour such as ear and tail chewing (Beattie *et al* 1998, p 27).

The first example statement (from Hill *et al* 1998) was linked to the assessment criteria 'biting' and 'AMI'. With respect to 'biting' (ie biteability of EMats) it specified that object dimensions and the size of the pigs were relevant. The statement also loaded on 'biting' through the Wcat 'preference' with a Wcat level score of 1, while it loaded on 'AMI' through 'preference' with a score of 2. For both scores the certainty code was specified as 'ref, exp', indicating that the statement contained a reference and reported an experiment.

The second example statement (from Beattie *et al* 1998) had the same certainty code but loaded on six different assessment criteria, namely 'rooting', 'exploration/learning', 'activity', 'AMI', 'tail biting' and 'penmate-directed behaviour'. On the first four criteria it loaded with a Wcat level score of 1 for 'preference', while it loaded on the two latter criteria ('tail-biting' and 'penmate-directed behaviour') with a score of –1 for the Wcat 'frustration/avoidance'.

The Wcat level scores of 1 (for positive welfare contributions) and –1 (for negative contributions) represent definite, but minimal loadings (as these scores range from 1 to 3, and –1 to –3, respectively). The higher score of 2 in the first example statement derived from the fact that there was not just a difference in manipulation between chain and hose, but there was a large difference. This score loaded on 'AMI' through 'preference', implying that the statement was interpreted as supporting the view that pigs prefer more, rather than less, interaction with an object such as a rubber hose (provided it was an appropriate size).

In this way, scientific statements provided arguments for 'adding' weights to assessment criteria. However, WFs were not derived by adding all Wcat level scores as this could result in double counting. Instead, the maximum Wcat level scores for each Wcat assigned to an assessment criterion were selected and these were added to generate WFs. This could mean that a single extreme finding could outweigh a body of more moderate responses, unless the single finding was classified as 'reduced certainty'. The certainty classification was used to correct the WF loadings following the principle that uncertain statements (as defined in the next section) resulted in a downward correction of the scores. In this way, design criteria were weighted, based on what was known about their relationship to performance criteria, while performance criteria were weighted based on what was known about the intensity, duration and incidence of the underlying emotions (cf Willeberg 1991; Anonymous 2001a).

In total, 679 different statements were transcribed into RICHPIG's database. These were analysed into 3,437 records specifying some relevant property for assessing enrichment. Out of these, 1,759 records were relevant for calculating WFs by specifying a relationship between an assessment criterion, a Wcat and a Wcat level score. For these 1,759 records, 573 statements were used. The other 1,678 records and the other 106 statements were used for other purposes, eg to further specify the meaning of the assessment criteria (without loading on their WFs), to specify (classes of) EMats and to specify background conditions (eg about the approach, the reference pen and the reference animal).

Out of 311 logically possible combinations of an assessment criterion and a Wcat (30 × 11) only 130 were actually used for weighting because they had received a Wcat level score and a certainty classification. The other logically-possible combinations generated a score of 0 by default.

'Preference' was by far the most frequently used Wcat (1,020 records), followed by 'frustration/aversion' (335 records) and 'natural behaviour' (122 records). Least frequently used were the Wcats 'survival' (10 records) and 'illness' (13 records). The frequent occurrence of 'preference' is not surprising as many scientific statements concerning EMats for pigs indicate a (relative) level of occupation, ie preference, of the EMat.

The most frequently encountered assessment criteria were animal-material interactions ('AMI', 369 records), 'rooting' (n = 142), 'explore/learn' (n = 141), 'tail biting' (n = 139) and 'chewing' (n = 122). The least frequently encountered criteria were 'sound' (n = 8) and 'disturbing' (n = 9).

Even though WFs were not directly derived from the number of statements or the number of relationships (records in the database), the Pearson's correlation coefficient between the number of relationships and the calculated WFs was reasonably high ($r = 0.83$, $P < 0.001$), while only one clear outlying value was observed in the scatterplot, namely for the assessment criterion 'AMI', where the number of relationships overestimated its calculated WF. This indicates that, in general, but not always, the more had been published about an assessment criterion, the more important it turned out to be, using the model's algorithm (which will be described more formally in the following section).

Calculation rules

For calculating WFs, the maximum Wcat level scores per Wcat were selected per assessment criterion. From this maximum Wcat level score, a score between 0 and 1 was subtracted depending on the level of uncertainty of the analysed relationship. This procedure was an attempt to build differences in certainty between different statements into the WF calculation procedure. The calculation of the certainty score (0 for most certain statements and 1 for least certain statements) was derived from the certainty classification identifying different types of arguments concerning the reliability of the scientific statements. This involved information about the number of citations (which was called the 'reference class'), about (the level of) 'scientific proof' (labelled 'observation class'), about the relevance of the subject matter, ie welfare effects of EMats for intensively-farmed pigs (labelled 'subject class'), and the quality of the paper in which the statement was retrieved (labelled 'paper class').

For example, the statement "movement and exploration have been suggested to be behavioural needs" (Fraser 1988) includes the term 'needs', which refers to a strong welfare claim. At the same time the statement says 'it has been suggested that', indicating that it is still only a hypothesis (and not 'scientific proof'). This 'discounted' some of the importance generated by the term 'need'.

The certainty score was set at 1 as a maximum, because subtracting 1 point from a Wcat level score of 1 (which was the most frequently assigned score on the 3-point scale) would reduce the score to a limit value of 0 (no effect) and would reduce Wcat level scores of 2 or 3 to 1 and 2, respectively. This seemed reasonable as it equated a certain Wcat level score with a highly uncertain Wcat subscore one level up.

The maximum weighting scores corrected for certainty scores were then 'added' into the WF for the assessment criterion.

More formally, the weighting factor (WF_i) of the i-th criterion in the model was calculated as:

$$WF_i = \left(\sum_{Wcat>0} \left(\underset{perWcat}{Max} \left(WcatLevSc - CerSc \right) \right) \right) - \left(\sum_{wcat<0} \left(\underset{perWcat}{Min} \left(WcatLevSc + CerSc \right) \right) \right)$$

Where *WF*: weighting factor (scale 0–33); *WcatLevSc*: weighting category level score (scale 1–3); *CerSc*: certainty score; positive Wcats (Wcat > 0) were 'natural behaviour', 'preferences' and 'demand', while the negative Wcats (Wcat < 0) were 'survival', 'illness', 'pain', 'stress', 'fitness', 'aggression', 'abnormal behaviour' and 'frustration/aversion'.

Certainty scores (CerSc, scale 0–1) were calculated as follows:

CerSc = ((([RC] + [OC] + [SC] + [PC])/4)–1)/(3.5–1)

Where, RC: 'Reference class', in which 1 point was assigned to 'many references cited', 2 points were assigned to 'more than one reference', 3 points were assigned to 'one reference' and 4 points were assigned to 'any statement without references' (which was the default). Note: the direction of the scale was adjusted to the principle that more points indicated less certainty. OC: 'Observation class', in which 1 point was assigned to an 'experiment', 2 points to a 'consensus statement', 3 points to a 'casual observation', 4 points to a 'hypothesis' and/or to a 'default statement' and 5 points were assigned to a 'statement classified as 'MB', ie a statement decomposition according to the scientific opinion of the modeller (ie MB). SC: 'Subject class', in which 1 point was assigned to 'intensively-farmed pigs (and default)', 2 points to 'other category of pig' or 'other reference pen' or to statements that applied 'across species', and 3 points were assigned to 'statements that both concerned other reference pens and another category of pig'.

PC: 'paper class', in which class A and class C papers received 1 point, while class B papers received 2 points.

To calculate the certainty score the sub-class points were added, divided by 4 (which was the number of sub-classes) and transformed to a scale from 0 to 1 (which required the '3.5' to be inserted into the formula as this happened to be the maximum sum of sub-class points in the present dataset).

The certainty scores were subtracted from absolute maxima of the Wcat level scores per Wcat, and the resulting scores were added to calculate the WFs, which were subsequently used to calculate enrichment scores.

This procedure deviates considerably from the conventional statistical expressions of error and uncertainty. The main reason for this was that the modelling focused on a pragmatic attempt to capture a biological reasoning process and that the information required for making these more formal calculations is lacking at present.

Formally, the overall enrichment score for an EMat treatment (ES_{EMat}, on scale 0 to 10) was calculated in the model as weighted averages, ie as the sum of the assessment criteria scores (*ACS*) multiplied by the weighting factor (*WF*) of each assessment criterion in the model, divided by their total sum of WFs:

$$ES_{Emat} = \frac{\sum_{i=1}^{m} (ACS_i \cdot WF_i)}{\sum_{i=1}^{m} WF_i}$$

Where ACS_i is the assessment criterion score (between 0 and 10) representing the degree to which the enrichment material (EMat) satisfies assessment criterion i; WF_i is the weighting factor of the i-th assessment criterion; m is the total number of assessment criteria in the model ie 30; $i \in [1, 30]$.

Some additional 'validation' results

Validation of a (semantic) model can take 3 forms: comparison with expert opinion; empirical validation and (sensitivity) analysis (Bracke *et al* 2002a). Validation of enrichment scores and weighting factors in relation to expert opinion was reported in two papers (Bracke *et al* 2007a, b). In these studies, some overlap between the experts (senior applied ethologists) and authors of publications used for modelling was unavoidable but this was not considered unacceptable because expert opinion was expressed at a group level. In another paper, a preliminary empirical study was reported (Bracke 2007) where support was found for 2 criteria in the model ('change/destructibility' and 'hygiene'), but less so for a third, less important criterion ('sound'). In addition, the absence of significant interactions in this study, as well as in a second study (Bracke & Spoolder 2008), was interpreted as providing some support for the assumption that (for the time being) interactions may be ignored for modelling. The results of a third empirical validation study, using an existing dataset from a study by van de Weerd *et al* (2003b) and the results of (sensitivity) analysis on robustness of the model, will be reported here.

The procedure of selecting maxima per Wcat for the calculation of weighting factors ensured that only the strongest statements counted within each subdomain (Wcat). This procedure was designed to avoid double counting of arguments that supported the weighting. The 'double' and less strong statements in the database, however, are not redundant as they provided a kind of robustness to the RICHPIG model. When the (analysis of the) stronger statements would have to be disqualified (based on some kind of criticism), these 'backup' statements could 'take over', limiting the breakdown of the model.

Out of the total of 130 primary weighting records (which were used for actual weighting) 18 records were backed up by at least 1 other record/statement (ie a scientific statement linked to an assessment criterion, Wcat, Wcat level score and certainty classification) such that deleting the primary record would not change the calculated WF. In total, 77 records were backed up by a statement that would result in a reduced WF when the primary statement would be deleted (median reduction: 0.2, range: 0.05 to 2.1). The remaining 35 records were not backed up (resulting in a median score reduction of 0.5 (range: 0.3 to 1.4).

When all primary records (ie those statements' decompositions used for the calculation of WFs in RICHPIG) were deleted, then (a) the median decrease in WF was 1.2 (range: 0.2 [for 'palatability']–4.4 [for 'explore/learn']) and (b) the Pearson's correlation coefficient of the two sets of WFs (for all 30 criteria) was 0.96 (which is very high; $P < 0.05$).

The Pearson's correlation coefficient between the WFs with and without correction for (un)certainty of the underlying scientific statements was also very high, namely 0.99 ($P < 0.05$).

A final indication of robustness of the model presented here is that when all WFs were set at 1, then the correlation coefficient of EMat scores calculated with and 'without' WFs was again very high (namely: $r = 0.99$, $P < 0.05$). This indicates that WFs may be considered to be redundant from a statistical (but perhaps not from a [bio]logical) point of view.

An empirical validation study involved the dataset from van de Weerd *et al* (2003b). These authors supplied 74 different EMats to 222 groups of 3 weaned pigs and 222 groups of 3 growing pigs. They measured the total interaction time with the materials (AMI) on day 1 and day 5 using focal sampling, and used a Multiple Stepwise Regression analysis (MSR) to determine which of 28 descriptors characterising the materials (using a 1–0 scale) most affected AMI.

A multiple linear regression analysis on all the characteristics showed that the maximum amount of variation (R^2) that could be explained for day 5 was 35.9% and for day 1 R^2 was 35.3%. MSR generated R^2 values of 25 and 29% for days 1 and 5, respectively.

The overall RICHPIG enrichment scores for these EMats were compared with the AMI values on days 1 and 5 using a General Linear Model (REML) and cube-root transformation with Genstat 7 (Anonymous 1993).

The R^2 of the RICHPIG scores to explain AMI measured on days 1 and 5 were 26 and 53%, respectively. The higher value for day 5 compared to day 1 indicates that RICHPIG is better able at assessing the longer-term enrichment for which it was designed. The score for day 5 was moderately high (53%) and slightly higher than the R^2's obtained by MSR, indicating that both methods are only moderately suitable to predict AMI. From the modelling perspective, this may be explained by the fact that RICHPIG was designed to assess overall enrichment value, in which AMI was regarded as only one component. In this context, it may be noteworthy that the R^2 of the AMI values of days 1 and 5 was 48%, indicating that AMI on day 1 is not a very good predictor of day 5 either (which is not surprising since EMats may change considerably over time due to the interaction of the pigs with the EMats). Finally, it was noted that MSR only partially resulted in the selection of descriptors that were also important in RICHPIG (especially 'ingestible' and 'destructible/change'). Most notably, MSR indicated that being rootable had a negative relationship with AMI (on both days 1 and 5), whereas in RICHPIG rootability was positively related to enrichment. This was caused by the fact that in the study by van de Weerd *et al* (2003b) hanging, ingestible objects (that were very popular with the pigs, eg carrots) were not being scored as 'rootable'. As van de Weerd *et al* (2003b) noted in their discussion, the positive relationship between rootability and enrichment is more in accordance with the current views of applied ethologists.

Discussion and conclusions

The objective of this paper was to explain RICHPIG's structure, including its underlying principles, conceptual framework, procedures, calculation rules and concepts, such as assessment criteria, weighting factors (WF) and weighting categories (Wcat). The paper should make it possible for critical evaluation to occur as well as helping to identify scope for further improvement and implementation in (policy-making) practice.

RICHPIG is a semantic model to calculate enrichment scores (scale 0 to 10) for enrichment materials (EMats) for intensively-housed, weaned growing and fattening pigs. The model contains 30 assessment criteria that were weighted based on a formal analysis of scientific information collected in the database.

The modelling methodology was derived from previous work (Bracke 2001; Bracke *et al* 2004a, b) and had several new features, including an enhanced implementation of the underlying biological assessment framework and an

https://doi.org/10.1017/S0962728600032218 Published online by Cambridge University Press

incorporation of a measure of uncertainty in the assessment algorithm. Both innovations were intended to address frequently received questions from scientists on semantic modelling. Neither, however, pretends to deliver definite answers. The present description gives an illustration of how enrichment materials for pigs can be assessed formally, based on available scientific information. The model is not 'finished' in that it did not prove feasible to incorporate the latest publications on enrichment materials for pigs, nor is the model designed, at present, for addressing individual- and time-dependent variation in great detail (ie the model applies more to 'pigs in general'). This indicates an important drawback: modelling takes time, is dependent on the state of the art in science, and may require periodic updating with new knowledge (depending on the rate of scientific progress).

The methodology and computer-based environment support the transparency of the model, allowing criticism of underlying principles and concrete choices made during modelling. An important advantage of this kind of modelling is that different points of view, eg with respect to the use of WFs, can be compared quantitatively. In the case of weightings, this resulted in the perhaps somewhat surprising result that the model is fairly robust. As with previous semantic models, RICHPIG is constructed so as to allow upgrades with new scientific information. This applies not only to new studies on EMats for pigs, but also to new ways in which the reasoning process from scientific 'facts'/findings to welfare conclusions can be represented.

At present the empirical support for the model is moderate and the congruence with expert opinion is reasonable. This may be taken to imply that the modelling could, in theory, be used to search for gaps in knowledge and improved fact finding (ie finding those facts that divert the most from current expert opinion). This would appear to be a sensible way forward for applied ethological science, resulting in testing predictions rather than merely comparing EMats.

RICHPIG was designed primarily to support the further implementation of EC Directive 2001/93/EC in the Netherlands. This directive states that: "Pigs must have permanent access to a sufficient quantity of material to enable proper investigation and manipulation activities, such as straw, hay, wood, sawdust, mushroom compost, peat or a mixture of such, which does not compromise the health of the animals". The modelling work generated scope for further implementation of this directive as it generated model scores and expert scores for a long list of EMats (which were mostly in accordance with the directive); (Bracke *et al* 2007a) and it generated a list of assessment criteria and their weighting factors (see also Bracke *et al* 2007b). The weighting factors can be used to select (the most important and feasible) assessment criteria which can be used to further specify the requirements for EMats for pigs. This latter approach could give room for innovations in farming practice because the model could allow prescribing policy targets in terms of assessment criteria and/or enrichment scores rather than in terms of policy means (such as a listing of EMats which are prescribed, eg straw, and which are forbidden, eg chains). Prescribing that EMats must, for example, be rootable or destructible, or that EMats must generate an overall score of at least 5 on a scale from 0 to 10, could give farmers the opportunity to enrich their pens with farm-specific solutions. It could also motivate them to design new and (ultimately) better alternatives, when they know what aspects are important and when they can anticipate the level of improvements made by such innovations. The benchmark materials assessed in RICHPIG provide an easy starting point for designing and evaluating new enrichment materials. RICHPIG has been used to evaluate welfare impact in relation to other values (such as economics and food safety issues); it has been used to formulate a policy statement regarding enrichment materials in the Netherlands and it has been used to design new rooting equipment which is presently being tested. Further application of semantic modelling in support of political decision-making, concerns the input provided by semantic modelling methodology and previous models (SOWEL and PIGTAIL) in European Food Safety Authority (EFSA) working groups addressing risk assessment in general (Bracke *et al*, a, b in press) and for pig welfare (EFSA 2007) in particular.

A major criticism of semantic modelling is that it is believed to be subjective. This objection is valid in that it is ultimately based on an interpretation of the meaning of scientific information and this implies, for example, that the quality of the model will depend on the quality of the information, the modeller and the user. However, the information and the procedures used should themselves be 'objective', ie based on valid science and the procedures are designed to take the modeller's point of view, as much as possible, out of the equation. In fact, the model makes it possible to quantify the degree of similarity. For example, a correlation of 0.92 ($P < 0.05$) was found between the model scores and the subjective scores given by the modeller (MB) for the EMats submitted to the experts (Bracke *et al* 2007a). In addition, the model contains prescriptions to limit the impact of the subjectivity of the user, eg that criteria scores for benchmark EMats must be used and that the integrity of the whole system must be maintained. This latter requirement also applies to the modelling and upgrading activities. For example, when one statement would be criticised effectively, this would not only lead to a disqualification of that statement, but it would require a critical re-evaluation of all other statement decompositions to which the same argument applies. This is necessary to maintain the internal consistency and reliability of the model.

Several areas of further work in semantic modelling can be identified. The model can (always) be upgraded with new information (following the rules described in this paper for modelling) and more validation studies can be designed to test (and upgrade) the model. In addition, the underlying modelling methodology could be applied to other areas, eg other species and (integration with) other values such as environmental, economic and fair trade concerns (see Bracke *et al* 2008). When a balance is maintained between feasibility on the one hand and validity (and reliability) on the other, semantic modelling activities can be used to organise the (scientific) facts into quantified assessments that can support political and ethical decision-making.

Acknowledgements

This work was funded by the Dutch Ministry of Agriculture, Nature and Food Quality (Animal Welfare programme). Heleen van de Weerd and co-authors and sponsors are greatly acknowledged for supplying the dataset from project AW1024 funded in the UK by Defra. Comments on earlier drafts were gratefully received from Sandra Edwards, Eberhard von Borell and Patrice Perny. Many thanks also to Johan Zonderland, Petra Lenskens and Edwin Bleumer for support in the database work.

References

Anonymous 1993 *Genstat 5 Release 3 Reference Manual, 1993*. Oxford University Press: Oxford, UK

Anonymous 2001a Scientists' assessment of the impact of housing and management on animal welfare. *Journal of Applied Animal Welfare Science 4*: 3-52

Anonymous 2001b *SPSS® Base 11.0 User's Guide*. SPSS: Chicago, Illinois, USA

Apple JK and Craig JV 1992 The influence of pen size on toy preference of growing pigs. *Applied Animal Behaviour Science 35*: 149-155

Arey DS and Maw SJ 1995 Food substrates as environmental enrichment for pigs. *Farm Building Progress 118*: 9-12

Beattie VE, Sneddon IA and Walker N 1993 Behaviour and productivity of the domestic pig in barren and enriched environments. Livestock Environment IV. *Fourth International Symposium* pp 42-50. American Society of Agricultural Engineers: St. Joseph, Michigan, USA

Beattie VE, Walker N and Sneddon IA 1996 An investigation of the effect of environmental enrichment and space allowance on the behaviour and production of growing pigs. *Applied Animal Behaviour Science 48*: 151-158

Beattie VE, Walker N and Sneddon IA 1998 Preference testing of substrates by growing pigs. *Animal Welfare 7*: 27-34

Beattie VE, Sneddon IA, Walker N and Weatherup RN 2001 Environmental enrichment of intensive pig housing using spent mushroom compost. *Animal Science 72*: 35-42

Blackshaw JK, Thomas FJ and Lee J-A 1997 The effect of a fixed or free toy on the growth rate and aggressive behaviour of weaned pigs and the influence of hierarchy on initial investigation of the toys. *Applied Animal Behaviour Science 53*: 203-212

Bøe KE 1992 The effect of different kinds of bedding on the behaviour of fattening pigs. *Proceedings of the CIGR Congress* pp 76-83. 8 September 1992, Szklarska Poreba, Poland

Bolhuis JE, Parmentier HK, Schouten WGP, Schrama JW and Wiegant VA 2003 Effects of housing and individual coping characteristics on immune responses of pigs. *Physiology and Behavior 79*: 289-296

Bracke MBM, Spruijt BM and Metz JHM 1999a Overall welfare assessment reviewed. Part 1: Is it possible? *Netherland Journal of Agricultural Science 47*: 279-291

Bracke MBM, Metz JHM and Spruijt BM 1999b Overall welfare reviewed. Part 2: Assessment tables and schemes. *Netherland Journal of Agricultural Science 47*: 293-305

Bracke MBM, Spruijt BM and Metz JHM 1999c Overall welfare reviewed. Part 3: Welfare assessment based on needs and supported by expert opinion. *Netherland Journal of Agricultural Science 47*: 307-322

Bracke MBM 2001 *Modelling of animal welfare: the development of a decision support system to assess the welfare status of pregnant sows*. PhD Thesis, Wageningen University, Wageningen, The Netherlands

Bracke MBM, Spruijt BM, Metz JHM and Schouten WGP 2002a Decision support system for overall welfare assessment in pregnant sows A: Model structure and weighting procedure. *Journal of Animal Science 8*: 1819-1834

Bracke MBM, Metz JHM, Spruijt BM and Schouten WGP 2002b Decision support system for overall welfare assessment in pregnant sows B: Validation by expert opinion. *Journal of Animal Science 8*: 1835-1845

Bracke MBM, Hulsegge B, Keeling L and Blokhuis HJ 2004a Decision support system with semantic model to assess the risk of tail biting in pigs: 1: Modelling. *Applied Animal Behaviour Science 87*: 31-44

Bracke MBM, Hulsegge B, Keeling L and Blokhuis HJ 2004b Decision support system with semantic model to assess the risk of tail biting in pigs: 2; 'Validation'. *Applied Animal Behaviour Science 87*: 45-54

Bracke MBM 2006 Expert opinion regarding environmental enrichment materials for pigs. *Animal Welfare 15*: 67-70

Bracke MBM and Hopster H 2006 Assessing the importance of natural behavior for animal welfare. *Journal of Agricultural and Environmental Ethics 19*: 77-89

Bracke MBM, Zonderland JJ, Lenskens P, Schouten WGP, Vermeer H, Spoolder HAM, Hendriks HJM and Hopster H 2006 Formalised review of environmental enrichment for pigs in relation to political decision making. *Applied Animal Behaviour Science 98*: 165-182

Bracke MBM, Zonderland JJ and Bleumer EJB 2007a Expert judgement on enrichment materials for pigs validates preliminary RICHPIG Model. *Applied Animal Behaviour Science 104*: 1-13

Bracke MBM, Zonderland JJ, Bleumer EJB 2007b Expert consultation on weighting factors of criteria for assessing environmental enrichment materials for pigs. *Applied Animal Behaviour Science 104*: 14-23

Bracke MBM 2007 Multifactorial testing of enrichment criteria: pigs 'demand' hygiene and destructibility more than sound. *Applied Animal Behaviour Science 107*: 208-232

Bracke MBM and Spoolder HAM 2008 Novel object test can detect marginal differences in environmental enrichment in pigs. *Applied Animal Behaviour Science 109*: 39-49

Bracke MBM, de Bakker E, Beekman V, Jansson K and de Graaff 2008 Voedselkwaliteit op het spoor – Op weg naar een transparant beleidsafwegingskader. *LEI Report 2008*. LEI: The Netherlands. [Title translation: Food quality on track towards a tranparent policy decision framework]

Bracke MBM, Edwards SA, Metz JHM, Noordhuizen JPTM and Algers B Synthesis of semantic modelling and risk analysis methodology applied to animal welfare. *Animal*, in press, a

Bracke MBM, Edwards SA, Engel B, Buist WG and Algers B Expert opinion as 'validation' of risk assessment applied to calf welfare. *Acta Veterinaria Scandinavica*, at press, b

Buré RG, Kerk Pvd and Koomans P 1983 *Het verstrekken van stro, compost en tuinaarde aan mestvarkens*: *Report 190*. IMAG: Wageningen, The Netherlands [Title translation: Providing straw, compost and garden mould to fattening pigs]

Courboulay V, Roubelet A and Loiseau D 2004 How do ground fixed or suspended objects influence the activity of growing-finishing pigs housed on fully slatted floor? *Journee Recherche Porcine 36*: 389-394

Day JEL, Spoolder HAM and Edwards SA 2001 Straw as environmental enrichment: which properties do growing pigs find behaviourally rewarding? *Proceedings of the International Symposium of the C.IGR* pp 157-166. 23-25 October 2001, Szklarska Poreba, Poland

Day JEL, Burfoot A, Docking CM, Whittaker X, Spoolder HAM and Edwards SA 2002a The effects of prior experience of straw and the level of straw provision on the behaviour of growing pigs. *Applied Animal Behaviour Science 76*: 189-202

Day JEL, Spoolder HAM, Burfoot A, Chamberlain HL and Edwards SA 2002b The separate and interactive effects of handling and environmental enrichment on the behaviour and welfare of growing pigs. *Applied Animal Behaviour Science 75*: 177-192

De Mol RM, Schouten WGP, Evers E, Drost WC, Houwers HWJ and Smits AC 2004 *Integrale welzijnsbeoordeling leghennen. Report 239*. Agrotechnology & Food Innovations of Wageningen University and Research Centre: Wageningen, The Netherlands. [Title translation: Integrated welfare assessment laying hens]

De Mol RM, Schouten WGP, Evers E, Drost WC, Houwers HWJ and Smits AC 2006 A computer model for welfare assessment of husbandry systems for laying hens. *Netherlands Journal of Agricultural Science 54*: 157-168

EFSA 2007 Scientific report on the risks associated with tail biting in pigs and possible means to reduce the need for tail docking considering the different housing and husbandry systems. *The EFSA Journal 611*: 2-98. Available at http://www.efsa.europe.eu/EFSA/efsa_locale-1178620753812_1178672658201.htm

Feddes JJR and Fraser D 1994 Non-nutritive chewing by pigs: implications for tail-biting and behavioral enrichment. *Transactions of the ASAE 37*: 947-950

Fraser D 1987 Mineral-deficient diets and the pig's attraction to blood: implications for tail-biting. *Canadian Journal of Animal Science 6*: 909-918

Fraser 1988 Behavioural needs in relation to livestock maintenance. *Applied Animal Behaviour Science 19*: 368-377

Fraser D, Phillips PA, Thompson BK and Tennessen T 1991 Effect of straw on the behaviour of growing pigs. *Applied Animal Behaviour Science 30*: 307-318

Gonyou HW and Bench CJ 2002 Effects of environmental enrichment at two phases of development on the incidence of belly nosing behaviour in early weaned pigs. In: Koene P (ed) *Proceedings of the 36th International Congress of the ISAE* p 114. 6-10 August 2002, Egmond aan Zee, The Netherlands

Grandin T and Curtis SE 1984 Toy preferences in young pigs. *Journal of Animal Science 59(S1)*: 85

Grandin T, Curtis SE and Taylor IA 1987 Toys, mingling and driving reduce excitability in pigs. *Journal of Animal Science 65(S1)*: 230-231

Heizmann V, Hauser C and Mann M 1988 Zum Erkundungs- und Spielverhalten juveniler Hausschweine in der Stallhaltung. *KTBL Schrift 323*: 243-265. [Title translation: On exploration and play behaviour of juvenile domestic pigs housed indoors]

Hill JD, McGlone JJ, Fullwood SD and Miller MF 1998 Environmental enrichment influences of pig behavior, performance and meat quality. *Applied Animal Behaviour Science 57*: 51-68

Höges J 1991 Beschäftigungsmaterial für Schweine. *DGS Magazin 17*: 492. [Title translation: Occupational materials for pigs]

Horrell I and Ness PA 1995 Enrichment satisfying specific behavioural needs in early-weaned pigs. *Applied Animal Behaviour Science 44*: 264

Jorgensen B 2003 Influence of floor type and stocking density on leg weakness, osteochondrosis and claw disorders in slaughter pigs. *Animal Science 77*: 439-449

Kress B, Ross A, van den Weghe H and Steffens G 1999 Straw as a material for occupying fattening pigs in a fully slatted floor system taking animal behaviour into special consideration. In: *Internationale tagung "Bau, Technik und Umwelt in der landwirtschaftlichen Nutztierhaltung"* pp 279-283. 9-10 March 1999, Freising, Germany

Krötzl H, Sciarra C and Troxler J 1993 Der Einfluss von Rauhfutterautomaten, Strohraufen und Nagebalken auf das Verhalten von Mastschweinen. *KTBL Schrift 361*: 181-191. [Title translation: Effects of roughage automats, straw racks and gnawing timber on the behaviour of fattening pigs]

Lyons CAP, Bruce JM, Fowler VR and English PR 1995 A comparison of productivity and welfare of growing pigs in four intensive systems. *Livestock Production Science 43*: 265-274

McKinnon AJ, Edwards SA, Stephens DB and Walters DE 1989 Behaviour of groups of weaner pigs in three different housing systems. *British Veterinary Journal 145*: 367-372

Meyer B, Hagelso M, Hansen LL and Jeppesen LL 1984 Effect of environment and rank order on agonistic behaviour in pigs. *Proceedings of the International Congress on Applied Ethology in Farm Animals* pp 162-165. 4-6 August 1984, Kiel, Germany

Moore EA, Broom DM and Simmins PH 1994 Environmental enrichment in flatdeck accomodation for exploratory behaviour in early weaned piglets. *Applied Animal Behaviour Science 41*: 3-4

Olsen AW, Vestergaard EM and Dybkjaer L 2000 Roughage as additional rooting substrates for pigs. *Animal Science 70*: 451-456

Olsen AW, Simonsen HB and Dybkjaer L 2002 Effect of access to roughage and shelter on selected behavioural indicators of welfare in pigs housed in a complex environment. *Animal Welfare 11*: 75-87

Pearce GP, Paterson AM and Pearce AN 1989 The influence of pleasant and unpleasant handling and the provision of toys on the growth and behaviour of male pigs. *Applied Animal Behaviour Science 23*: 27-37

Pearce GP and Paterson AM 1993 The effect of space restriction and provision of toys during rearing on the behaviour, productivity and physiology of male pigs. *Applied Animal Behaviour Science 36*: 11-28

Petersen V, Simonsen HB and Lawson LG 1995 The effect of environmental stimulation on the development of behaviour in pigs. *Applied Animal Behaviour Science 45*: 215-224

Sambraus HH and Kuchenhoff R 1992 The effects of environmental objects on the resting behaviour and behavioural abnormalities of piglets. *Tierarztliche Umschau 47*: 233-242

Sambrook TD and Buchanan-Smith HM 1997 Control and complexity in novel object enrichment. *Animal Welfare 6*: 207-216

Schaefer AL, Salomons MO, Tong AKW, Sather AP and Lepage P 1990 The effect of environment enrichment on aggression in newly weaned pigs. *Applied Animal Behaviour Science 27*: 41-52

Spoolder H, De Rosa G, Hörning B, Waiblinger S and Wemelsfelder F 2003 Integrating parameters to assess on-farm welfare. *Animal Welfare 12*: 529-534

Stubbe A 2000 *Entwicklung und Beurteilung einer Beschäftigungstechnik für Mastschweine in intensiven Haltungssystemen*. PhD thesis, Institut für Agrartechnik, Universität Hohenheim, Hohenheim, Germany. [Title translation: Development and assessment of an occupation technique for fattening pigs in intensive housing systems]

SVC (Scientific Veterinary Committee) 1997 The welfare of intensively kept pigs. *Report to the Directorate General XXIV of the European Commission*. Adopted 30th September 1997. Doc. XXIV/ScVc/0005/97, Scientific Veterinary Committee, Animal Welfare Section: Brussels, Belgium

van de Weerd HA, Docking CM, Day JEL, Breuer K and Edwards SA 2003a Longitudinal study of adverse behaviour of undocked pigs in two different housing systems. *Proceedings of the British Society of Animal Science.* 9-10 September 2003, Nottingham, UK

van de Weerd HA, Docking CM, Day JEL, Avery PJ and Edwards SA 2003b A systematic approach towards developing environmental enrichment for pigs. *Applied Animal Behaviour Science 84*: 101-118

van Putten G 1980 Objective observations on the behaviour of fattening pigs. *Animal Regulation Studies 3*: 105-118

van Rooijen J 1981 Die Anpassungsfähigkeit von Schweinen an einstreulose Buchten. *KTBL Schrift 281*: 174-185. [Title translation: Adaptability of pigs to strawless pens]

van Rooijen J 1982 Operant preference test with pigs. *Applied Animal Ethology 9*: 87-88

Wiepkema PR 1987 Behavioural aspects of stress. In: Wiepkema PR and van Adrichem PWM (eds) *Biology of Stress in Farm Animals: An Integrative Approach* pp 113-133. Martinus Nijhoff: Dordrecht, The Netherlands

Wiepkema PR and Koolhaas JM 1993 Stress and animal welfare. *Animal Welfare 2*: 195-218

Willeberg P 1991 Animal welfare studies: Epidemiological considerations. *Proceedings of the Society of Veterinary Epidemiology and Preventive Medicine* pp 76-82. 17-19 April 1991, London, UK

Wood-Gush DGM and Beilharz RG 1983 The enrichment of a bare environment for animals in confined conditions. *Applied Animal Ethology 10*: 209-217

Wood-Gush DGM and Vestergaard K 1991 The seeking of novelty and its relation to play. *Animal Behaviour 42*: 599-560

Zonderland JJ, Fillerup M, Van Reenen CG, Hopster H and Spoolder HAM 2003a *Preventie en behandeling van staartbijten bij gespeende biggen. RIAH report 18.* Praktijkonderzoek: Lelystad, The Netherlands. [Title translation: Prevention and treatment of tail biting in weaned piglets]

Zonderland JJ, Vermeer HM, Vereijken PFG and Spoolder HAM 2003b Measuring a pig's preference for suspended toys by using an automated recording technique. *CIGR Ejournal V*: 1-11