

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Massey University Library

Thesis Copyright Form

Title of thesis: *The... Acquisition of Knowledge from Multiple Experts in the Domain of Sensory Evaluation Panel Training*

(1) (a) I give permission for my thesis to be made available to readers in Massey University Library under conditions determined by the Librarian.

~~(b) I do not wish my thesis to be made available to readers without my written consent for months.~~

(2) (a) I agree that my thesis, or a copy, may be sent to another institution under conditions determined by the Librarian.

~~(b) I do not wish my thesis, or a copy, to be sent to another institution without my written consent for months.~~

(3) (a) I agree that my thesis may be copied for Library use.

~~(b) I do not wish my thesis to be copied for Library use for months.~~

Signed *[Signature]*

Date *15. Sep. 94*

The copyright of this thesis belongs to the author. Readers must sign their name in the space below to show that they recognise this. They are asked to add their permanent address.

NAME and ADDRESS

DATE

**THE ACQUISITION OF KNOWLEDGE FROM MULTIPLE EXPERTS
IN THE DOMAIN OF SENSORY EVALUATION PANEL TRAINING**

A Thesis Presented in Partial Fulfilment
of the Requirements for the
Degree of Master of Philosophy in Social Science
(Computer Science)
at Massey University, New Zealand

SIMON ANTONY EWING-JARVIE

1994

Acknowledgements

I would like to express my sincere thanks to my supervisor, Elizabeth Kemp. Not only for her support and guidance throughout the project but particularly for her grace in maintaining the delicate equilibrium between allowing me the freedom to study independently and still keeping me moving toward the objective. My thanks must also go to the many staff of the department of Human Resource Management at Massey University who assisted me with ideas on layout and with proof-reading.

A special mention must go to the staff of the Sensory Evaluation Unit at the New Zealand Dairy Research Institute, Palmerston North. In particular, I would like to thank Sharon Wards, for believing in the project at the outset, and also Pam Marks and Anne McPherson. Without their time and interest the research contained in this thesis could not have been completed.

The drive to achieve in life is not innate and this thesis would not be complete without a thank you to my father who has set the example for life long learning from as far back as I can remember. I hope this work helps to motivate my children, Timothy and Amy, to achieve in the same way that Robert Ewing-Jarvie's unquenchable thirst for knowledge has inspired me. Thankfully, the children are too young to have noticed their father's odd behaviour during the two years that this research has been conducted.

Simon Ewing-Jarvie
September 1994

Abstract

Knowledge acquisition is the elicitation and representation of human expertise and is one of the first steps taken in constructing an expert system. It has often been cited as the 'bottleneck' in expert systems development due to the labour intensive processes needed to deal with the expert human. Various researchers have proposed methodologies for improving both the accuracy and the productivity of the process. This has ranged from manual to automated methods as well as examining what the expert might be thinking during a study of the conscious activity.

This research has focused on the issues involved in the manual elicitation of knowledge using multiple experts in the same domain. It utilises the transcripts of semi-structured interviews and discourse analysis techniques to construct the domain layer of a knowledge base, employing the KADS methodology.

The findings highlight the importance of the relationship between the knowledge engineer, organisation and the human experts. Issues such as motivation, organisational commitment and communication skills feature as key indicators of the likely success of an expert system development project. While automated acquisition assists with productivity, it works against the development of relationships within the project team and the trade-off must be carefully considered by the project manager.

Table of Contents

1 Introduction	1
1.1 An overview	1
1.2 Nature of the research	1
1.3 Structure of the thesis	2
2 Knowledge Acquisition	4
2.1 General	4
2.2 Knowledge acquisition techniques	6
2.2.1 Interviewing	6
2.2.2 Psychologically-based techniques	7
2.2.3 Secondary data	9
2.3 Models of problem solving	9
2.3.1 Method to task approaches	9
2.3.2 Generic task approaches	12
2.3.3 The KADS approach	12
2.3.4 Role limiting approach	13
2.3.5 Second generation techniques	14
2.3.6 Future generations	14
2.4 Criteria for selecting experts	16
2.5 Use of multiple experts	17
2.5.1 General	17
2.5.2 Need for multiple experts	17
2.5.3 Examples of the use of multiple experts	17
2.5.4 Problems of multiple expertise sources	18
2.6 Conclusion	19
3 Sensory Evaluation Panels	20
3.1 Introduction	20
3.2 Sensory evaluation defined	20
3.3 Selecting panellists	21
3.4 Training panellists	21
3.5 Panel leaders	22
3.6 Statistical aspects of panels	22
3.7 Suitability as a domain	22
4 Methodology	24
4.1 General	24
4.2 Detailed methodology	24
4.2.1 The domain	24
4.2.2 The experts	25
4.2.3 The interviews	26
4.2.4 Initial analysis	26
4.2.5 Further analysis	29
4.2.6 Development of consensus	32

4.3 Conclusion	33
5 Analysis	34
5.1 Interviews	34
5.2 Transcript analysis	34
5.2.1 Phase 'A' analysis	35
5.2.2 Phase 'B' analysis	35
5.3 Expert 1	35
5.4 Expert 2	37
5.5 Expert 3	39
5.6 Consensus vocabulary	41
5.7 Development of domain layer model	44
5.8 Conclusion	57
6 Discussion	59
6.1 Introduction	59
6.2 Key findings	59
6.3 Organizational commitment	60
6.4 Identification of experts	60
6.5 Research design	61
6.6 Organisational and individual commitment	62
6.7 Processing of interview data	62
6.8 Reduction of transcripts	63
6.9 Discourse analysis of individual transcripts	64
6.10 Achievement of consensus	64
6.11 Benefits to the experts	65
6.12 Benefits to the organisation	65
6.13 Conclusions	65
7 Conclusions	67
7.1 A modelling approach	67
7.1.1 The value of manual elicitation	67
7.2 Discourse analysis	67
7.3 Issues relating to multiple experts	68
7.3.1 Contamination potential	68
7.4 Organizational issues	69
7.5 Redefining the role of the knowledge engineer	69
7.6 Directions for future research	69
7.7 Closing summary	70
References	71
Appendices (with separate contents page and numbering)	79

Chapter 1

Introduction

1 Introduction

1.1 An overview

The field of artificial intelligence and its offspring, expert systems has been the subject of considerable interest in recent years. Various international initiatives have added to this impetus including Japan's Fifth Generation Strategy, the European Community's KADS Programme and the British Alvey Programme for Advanced Information technology. Expert systems utilise human expertise and there has been much research (Clancey, 1985; Boose & Gaines, 1989; Wielinga, Schreiber & Breuker, 1991) into knowledge engineering techniques. Despite the proliferation of research projects in the area, there is general agreement (Clancey, 1985; Marcus, 1988; Musen, 1989; Wielinga, Schreiber & Breuker, 1991) that the major bottleneck in the construction of an expert system remains the acquisition of knowledge. This problem is further confounded when multiple experts in the same domain are involved in the knowledge acquisition process. There is a limited amount of research on the acquisition of knowledge from multiple experts in the same domain.

1.2 Nature of the research

This research was exploratory and its objective was to identify issues that will affect knowledge acquisition from multiple experts in the same domain. While there are now second and third generation software applications which facilitate the development of a knowledge base (Marcus, 1988; Musen, 1989; Gaines & Linster, 1990), it was felt that too little was understood about some basic relationships between knowledge engineer, expert and domain and that automation would blur the issues. For this reason, first generation techniques such as interviewing and discourse analysis were employed. This proved to be a valuable feature of the research as it enabled the effect of the relationship between knowledge engineer and expert to be fully explored within the context of organizational dynamics.

1.3 Structure of the thesis

In the second chapter the literature on knowledge acquisition is reviewed and definitions are offered for the main terms and ideas implicit in the research. It is pertinent to note that most of the published research relates to knowledge acquisition involving a single expert, rather than multiple experts.

Expert systems are symbolic, heuristic, and can cope with uncertain information. The field of sensory evaluation is of a similar nature, being involved with the use of the human senses to evaluate flavour, aroma, texture, mouth feel and colour in a quantifiable, scientific manner. Members of a panel are trained and led by sensory evaluation experts in order to enable them to carry out these experiments. The Sensory Evaluation Unit of the New Zealand Dairy Research Institute, based in Palmerston North is one of the main providers of trained panels in the country and as such, was in the position of having three expert panel leaders on its staff. This is contrary to the normal situation where expertise is scarce or unavailable and so offered a unique opportunity for the study of knowledge acquisition using multiple experts. The third chapter provides an overview of the field of sensory evaluation as well as discussing its suitability as a domain.

Chapter four details the methodological processes employed in the research. As is the nature of exploratory research, a research goal was deemed more appropriate than hypotheses. This chapter describes how the verbal data was gathered and transcribed. It explains how original interview transcripts were reduced via discourse analysis techniques to eventually become concepts and relations. Despite the best efforts of researchers, this process will remain relative to the skill of the knowledge engineer and so consistency is only maintained, for the purposes of research by using one knowledge engineer, as is the case here. However, it was noted that the use of one knowledge engineer does result in a bottleneck of administrative tasks.

The concepts and relations derived from discourse analysis form the basis of the domain layer, in the KADS methodology. KADS is explained more fully in

chapter 2, while the concepts and relations are shown in chapter 5. The KADS methodology was selected from among several alternatives as it was clear that many of the tasks involved in training a sensory evaluation panel were generic, for instance selection, discrimination and administration. KADS offered the most clearly documented framework upon which to model these activities. As the DRI had requested that the outcome be a single model of knowledge, it was then necessary to achieve consensus and derive one set of concepts and relations from the three experts' individual sets. Chapter 5 also addresses the processes which led to the successful development of a single domain layer.

The significant issues identified in this research are examined in chapter 6. These findings relate particularly to social science concepts such as organizational and individual commitment, communication, motivation and workload. This chapter serves to highlight the inseparability of computer science from the disciplines which seek to explain what energises and sustains human behaviour.

Conclusions regarding the findings are highlighted in the final chapter, chapter seven. The verbal data transcripts and output from supporting analysis is shown in full in the appendices. Despite the volume of the data, it was felt that its inclusion not only provided completeness to this thesis, but also offered the opportunity for secondary analysis in the future.

Chapter 2

The Literature on Knowledge Acquisition

2 Knowledge Acquisition

2.1 General

Knowledge acquisition (KA) is the elicitation and representation of human expertise. In early research into the field it was seen as a preliminary activity to the construction of an expert system. This can be seen with such definitions as "the transfer and transformation of problem solving expertise from some knowledge source to a program" (Buchanan, Barstow, Bechtel, Bennet, Clancey, Kulikowski, Mitchell & Waterman, 1983). Later researchers found that it did not necessarily lead to the production of a computer program. Gammack (1987) considered KA had four distinct phases: eliciting concepts, eliciting structure between concepts, representing structure and finally transforming that structure for some desired purpose. KA is a facet of expert systems (ES) development but to properly understand its meaning, one must view it in its context as a part of the overall field of artificial intelligence (AI). AI is described by Barr & Feigenbaum as:

"the part of computer science concerned with designing intelligent computer systems, that is, systems that exhibit the characteristics we associate with intelligence in human behaviour - understanding language, learning, reasoning, solving problems, and so on" (1981, p1).

Following on from this definition, the distinct AI sub-disciplines of expert systems, robotics, and natural language processing and computer vision can be identified (Jackson, 1990). ES are also referred to as knowledge-based systems (KBS) which, like other AI systems, are symbolic, heuristic, inferential, and can cope with uncertain information.

"An expert system is a computer program that represents and reasons with knowledge of some specialist subject with a view to solving problems or giving advice" (Jackson, 1990, p3).

ES differ from other kinds of AI systems as they normally deal with large amounts of real human expertise and attempt to solve problems of genuine scientific or commercial interest.

Given that ES contain large amounts of human expertise, it stands to reason that there must be a process for acquiring, representing and validating this expertise. This is the field of knowledge engineering and, as the process suggests, encompasses knowledge acquisition. The term knowledge acquisition (KA) however, as Jackson (1990) noted, is generic and gives no insight into how expertise is transferred. Unlike classic systems development, KA is a cyclic rather than linear process. Cordingley (1989) notes that there is considerable inconsistency regarding the difference between KA and Knowledge elicitation (KEL). KEL often implies human interaction between knowledge engineer and expert, although it could be extended to include indirect interaction with experts in any systematic way, e.g. sample problem solution. Cordingley defines KEL as:

"those activities undertaken by a person, the knowledge elicitor, to obtain material from any relevant source, analyze and interpret that material, and put in a pre-encoded form which, while useful to those who will encode the knowledge in the KBS language, also allows it to be scrutinised by all parties interested in KBS development" (1989, p93).

KA processes have traditionally been very labour intensive and the relatively low output, in engineering terms, has led the ES community to refer to KA as the 'bottleneck problem' in applications development. Traditionally KA techniques have involved a knowledge engineer working within a team comprising one or more subject matter experts. In an attempt to solve the 'bottleneck', integrated knowledge acquisition tools have been developed which interview experts, model behaviour, analyze text, form conceptual structures or generate inference rules (Boose & Gaines, 1989). Other productivity gains have been made through the use of models or templates such as KADS (Akkermans, Van Harmelen, Schreiber & Wielinga, 1992). KADS is supplemented by a range of software support tools which will ultimately provide an integrated workbench to support the knowledge

engineer. These tools include analysis, wordprocessing, documentation and model libraries.

2.2 Knowledge acquisition techniques

2.2.1 Interviewing

Manual KA techniques such as interviewing are the most widely used and generally applicable techniques. They are usually able to be used in any domain and can often fill the gap left by later generation automated techniques (LaFrance, 1987). Methods of questioning are the most accepted form of interviewing and these are divided into structured, semi-structured and unstructured approaches. The level of structure is the degree to which the KE prepares the interview format in advance, with semi-structured interviews having a prepared list of questions but the order and vocabulary being modified by the KE to suit the situation (Cordingley, 1989). Unfocussed interviews are more common at the outset of a KA process in order to establish the context of a domain, with alternative techniques being required as the elicitation sessions progress and knowledge is extracted more slowly (Hart, 1992). Interview sessions are normally recorded on audio or video tape and subsequently transcribed. Other less comprehensive recording techniques include the use of field notes and interview record forms. This transcript is then examined using the techniques of protocol analysis discussed in the next section. Regardless of the level of structure employed within a session, the KE will choose either to focus on a specific aspect of the domain or to follow a system by which he or she believes all the relevant material will be elicited.

There are many inherent inaccuracies in interviewing. Musen (1989) notes that a myriad of well constructed questions may still elicit misinformation due to the fact that much expertise is stored in ways which are unavailable to the consciousness. Given that there are now established psychologically based methodologies for systematic analysis, the problems associated with the transfer of ideas have been largely overcome (Hart, 1992). Hart notes

that there are many ways of conducting an interview and that the term encompasses a range of elicitation techniques which are based in the discipline of psychology. Techniques such as LaFrance's (1987) knowledge-acquisition grid have been developed in order to make manual KA more systematic.

2.2.2 Psychologically-based techniques

Formal psychometric techniques developed in the discipline of cognitive psychology have become popular tools in efforts to overcome the tacit nature of expertise. The most widely studied methods in this category involve protocol analysis which is a technique involving the study of experts' dialogue while they are solving a real or simulated case. This style of analysis has a wide spectrum of designs. Ericsson and Simon (1984) discuss four methods: think aloud (relating internal thoughts, talk aloud (relating internal 'conversations'), eidetic reduction (observations on own behaviour) and retrospective reporting of an activity. The subsequent verbal protocol can then be transcribed and analyzed to build a model of the execution of a task (Newell & Simon, 1972, Ericsson & Simon, 1984; Belkin & Brooks, 1987; Regoczei & Plantinga, 1987, Kuipers and Kassirer, 1987). There are some criticisms of protocol analysis. The first three of Ericsson and Simon's methods are conducted in real time and the need to articulate thoughts may actually interfere with the performance of the task. There are doubts about the validity and reliability of verbal reports as data (Diaper, 1987) although Ericsson and Simon (1984) discuss this issue at length and conclude that there is value in the data.

Another widely used psychological technique has been the use of repertory grids and associated ideas from personal construct theory (Kelly, 1955). These techniques have made a particularly important contribution because of their inherent suitability for automation. The most common method of employing this approach is to construct triads or dyads between objects or concepts to see why a subject views them as similar or different. Manual

applications of this technique are time consuming and can be unnerving to a subject who is unsure as to what information is being elicited.

Variations on the use of the psychological principles underpinning grid techniques are found in card sorting and laddering. Sorting tasks can be conducted in a number of ways. One method involves the subject separating cards showing elements of the domain into two piles and labelling each as to their content. Each sub-pile is further divided giving a binary tree decomposition. Alternatively a subject can sort cards two at a time, putting them in piles on the basis of similarity or difference (Cordingley, 1989). Laddering (Burton, Shadbolt, Rugg & Hedgecock, 1988; Graham and Jones, 1988) is a technique considered valuable for establishing hierarchies of concepts. The KE employs 'WHY' questions to get superordinate concepts, 'HOW' questions for subordinate concepts, and 'GIVE ALTERNATIVE EXAMPLES OF' questions to get concepts at the same level. Subjects can become tired of continual questioning of this nature. The KE needs also to be sensitive to probing on core concepts which may place pressure on the subject's own self image or esteem, and alienate them from the project. Cooke and McDonald (1987) introduced the idea of multidimensional techniques including statistical approaches such as cluster analysis, believing that this offsets the bias inherent in introspection.

Because of the subconscious nature of much human expertise, it may be difficult to differentiate between causal and random relationships. Any cognitive task is subject to bias, particularly due to shortcuts allowed for in complex heuristics. Cleaves (1987) identified techniques to combat bias in KA such as visual props, response mode changes, scoring rules, consensus weighting, focusing, decomposition, training, logic challenges and consensus interaction.

2.2.3 Secondary data

Although expertise is normally elicited or modelled from a human source, it is also possible to conduct knowledge engineering activities on secondary data sources such as reports, statistics, job aids and regulations. Much knowledge is available within the physical or social environment of the domain. This can be gathered through a variety of techniques such as observation, object tracking or collection of artefacts. *FIRECODE* is an expert advisory system developed at the University of Auckland to help building designers conform to the very complex fire safety code for particular types of buildings. The primary knowledge source was not a human but the document DZ4226 Code of Practice for Design for Fire Safety (Buis, Hamer, Hosking, & Mugridge, 1986).

2.3 Models of problem solving

Due to the 'bottleneck' problem and also because of the enormous cost of single run KA projects, there has been increasing pressure to develop systems which can imitate or assist the expertise of a knowledge engineer. The major focus for this has been toward tools that can elicit, organise and use domain knowledge to perform a specific task and consequently such tools make strong assumptions about the problem solving method to employ (Marcus, 1988). Various researchers have attempted to categorise the types of approaches possible. Musen (1989) divided the range into representation, method or task-based models. Since this group of categories lacks full coverage of modelling approaches such as KADS the divisions, for the purpose of this study are those proposed by Karbach, Linster and Voss, (1990). They defined the approaches as either a method to task, generic task, model-based or role limiting conceptual approach.

2.3.1 Method to task approaches

In this category the content knowledge is modelled separately from the problem solving knowledge thereby facilitating development through separation of the tasks of the KE and expert. Early attempts at automating KA were based around the use of predicate logic and were implemented

by McCarthy (1968) in the *Advice Taker* application. This approach paved the way for later developments in knowledge representation. *Teiresias* (Davis, 1977), initially designed to acquire knowledge for the medical ES *Mycin* (Shortliffe, 1976) used a limited natural language interaction to build inference rules. Metalevel knowledge was derived automatically from syntax in this system which allowed for identification of missing or incorrect rules during construction.

While production rules are useful in representing the relationship between conditions and actions, some domains require a focus on the properties and interrelationships between objects. Several representation-based approaches have followed from *Teiresias*, including *KAS*, the KA tool for *Prospector* (Duda & Gaschnig, 1981). Representing the knowledge base as a semantic network of nodes rather than rules, *KAS* included two key components. One of these, *RENE* (REsident Network Editor), was used to construct the semantic net while the other, the *Semantic Network Matcher*, examined the relationship between facts and knowledge in the network. A further representation method is that of frame systems (Minsky, 1975). These use prototypical sets of knowledge which apply to the majority of instances of an object. As such, they can be considered complex data nodes within an associative net. Because of the lack of clarity regarding the boundaries of a frame, the potential for ambiguity or error through clashes from multiple inheritance exists.

Method-based KA tools focus on the knowledge level of a knowledge base and users of such tools employ symbols to describe behaviours that the final ES must achieve (Musen, 1989). Dialogue between software and human does not centre on rules or frames but on the manner in which the problem is solved. An example of this approach is found in Clancey (1985), in an examination of *Heracles* (Heuristic Classification Shell)(Wilkins, Clancey & Buchanan, 1987), which was developed in order to understand

how to generalise *Neomycin* (Clancey & Letsinger, 1981). Heuristics about KA were used by Lefkowitz & Lesser (1988) to assimilate expertise into an existing knowledge base.

Developers of medical ES have produced other KA tools such as *Roget* (Bennet, 1985) which was designed to assist KA for systems based on the *Emycin* (van Melle et al, 1980) model of problem solving. Kahn, Nowlan & McDermott (1985) developed a system called *More*, which used a similar problem solving approach to *Roget*. It used heuristics about KA to identify possible problems in the knowledge being entered such as instances where competing hypotheses lacked distinctiveness. It was difficult to use however, because of constant requests on the user to enhance the knowledge base. A system based on *More* which eliminates these difficulties is *Mole* (Eshelman, Ehret, McDermott & Tan, 1987) which employs questions regarding refinements only when errors are encountered in a consultation session.

The foregoing programs required the user to possess an understanding of the model in use. A departure from that theme is found in the systems based on personal construct theory (Kelly, 1955) which solicit knowledge while concealing the problem-solving method thereby spanning the area between manual and automated psychological approaches. *ETS* (Expertise Transfer System) conducts a structured interview and elicits the expert's personal constructs in an automated version of the original grid. Analysis of the grid provides implied hierarchies among constructs which can, via an ES shell program, be used to generate a knowledge base (Boose, 1985). It is not suitable for all domains, however as it provides an incomplete model of knowledge due to the limited and sometimes incomplete relationships revealed in a grid. It does provide a quick platform from which knowledge engineers can work using traditional techniques. A more formal methodology for this approach was developed by Garg-Janardan & Salvendy (1988).

A successor to *ETS*, known as *Aquinas*, is under development by Boose & Bradshaw (1987). This system allows users to specify multiple hierarchies but requires an understanding of the heuristic classification model of problem solving. Other repertory grid based systems include *Kitten* (Shaw & Gaines, 1987)

2.3.2 Generic task approaches

Generic tasks are basic combinations of knowledge structures and problem solving strategies (Bylander & Chandrasekaran, 1987). This approach works at identifying recurring types of reasoning performed by ES and developing a framework for their implementation. Applications falling into this category deal with predefined approaches to solving general tasks. The system elicits information from the user in order to be able to relate its generic model to a specific task (Musen, 1989). *Opal* (Musen, Fagan, Combs & Shortliffe, 1987) contains a generic set of terms and relationships relating to the treatment of cancer. It rests on the premise that all treatments are variations of what is currently possible and is used as the knowledge editor for *Oncocin* (Shortliffe, Scott, Bischoff, van Melle and Jacobs, 1981) an ES to assist physicians in planning cancer patient care.

Knack (Klinker, Bentolila, Genetet, Grimes & McDermott, 1987) enables domain experts to create expert systems that assist in reporting tasks. Its acquire-and-present paradigm is complementary to the approach of *Salt* (Marcus, 1988). *Salt* departs from the style of most other task-based tools by implementing a strategy of propose-and-revise problem solving. It constructs a knowledge-base incrementally by proposing values for parameters, identifying constraints on these and revising decisions by analysing violations of these constraints.

2.3.3 The KADS approach

KADS (Wielinga, Schreiber and Breuker, 1991) is a methodology for building knowledge-based systems (KBS) initiated in 1983 by the ESPRIT

Programme of the Commission of the European Communities. The KADS methodology views KBS development as a modelling activity. It is based on five principles which are:

- a. the use of abstract intermediate models which are free from issues of implementation.
- b. a four layer conceptual model encapsulating the strategic, task, inference, and domain layers of the KBS.
- c. the re-use of partial or whole models for generic problem solving tasks.
- d. developing simple models into more complex ones.
- e. the preservation of structure during design and implementation.

The use of the four layer conceptual model is helpful in dealing with the relative complexity of real world problems. The conceptual model's four layers contain discrete elements. The domain layer contains the static knowledge such as concepts and relations. The second layer, the inference layer, describes what inferences can be made but not how or when they are made. The task layer describes how knowledge sources can be combined to achieve a goal and is in effect, the control layer of the model. The strategic layer is less clearly defined but is designed to contain meta-rules, which, in the event of a problem, can be used to modify the task layer (Schreiber, Breuker, Bredeweg & Wielinga, 1988). This can be likened to an expert learning from experience. *KADS* appears to be the only current approach to KBS development that has the potential to become the comprehensive, structured approach needed for full commercial support of Expert Systems.

2.3.4 Role limiting approach

This approach is described by Karback, Linster & Voss (1990) as a specialization of a weak method. By weak method it is meant that the problem-solving method is only weakly constrained by task features. The method employs little task specific knowledge and employs an algorithm to specify what knowledge identify or select. This would typically be

implemented in a simple loop. A feature of this method is its broad applicability due to its lack of task dependence. However, there are many tasks that cannot be performed without specific control knowledge.

2.3.5 Second generation techniques

Early attempts at automated KA have been based around the use of a single tool which, of necessity, performs usually only one aspect of the acquisition process. Knowledge gleaned then had to be supplied to an ES shell to create a knowledge base. The second generation of these tools explores the overall aim of creating knowledge support systems which are capable of multi-functional support to the user. This is achieved through the loose coupling of different software tools via protocols so that, to the user, it appears that the software functions as one unit (Gaines & Linster, 1990).

An example of this concept is the coupling of *Hypercard*, *KSS0* and *Babylon*. *Hypercard* allows users to input unstructured knowledge such as diagrams, or to annotate other elicited knowledge. *KSS0* (Knowledge Support System Zero) is a KA tool which provides more structured elicitation. *Babylon* is the ES shell which performs inferences from the knowledge base. All three communicate, with *Hypercard* being the interface for the elicitation of knowledge during development or provision of data at runtime (ibid).

2.3.6 Future generations

Further development of automation for KA is progressing in two directions, purpose built knowledge engineering workbenches and machine learning. The emphasis in the former is on a coherent integration of a wide variety of the first generation techniques, together with other support tools such as documentation systems (Shadbolt & Wielinga, 1990). This concept is similar to the integrated project support environment (IPSE) approach advocated for conventional software engineering projects (Sommerville,

1989).

KEW (Knowledge Engineering Workbench) is one such approach. In order to achieve its aims, *KEW* is, in effect, a 'meta-KBS' possessing knowledge about KA (Shadbolt & Wielinga, 1990). It employs a set of generic models to assist the knowledge engineer and is also designed to provide clerical support, KA tool execution, knowledge transfer and transformation, and context sensitive advice.

Machine learning refers to the ability of a computer application to develop, maintain and enhance its own knowledge base, thereby removing or substantially reducing the labour intensive requirement for knowledge engineers (Shalin, Wisniewski, Levi & Scott, 1988). It is already in limited use in ES development through the induction of rules from data (Quinlan, Compton, Horn & Lazarus, 1986). Decision tree induction using domain knowledge is a variation on this approach (Nunez, 1990). The techniques are problematic however as it must be assumed that data are correct and examples are error-free. Missing or incorrect information is referred to as noise. This limits the reliability of the approach for commercial applications (Kodratoff & Manago, 1987; Pettit, 1987).

Another approach has been to analyze cases or analogies (Schmidt & Schmalhofer, 1990, Ruqian & Cungen, 1990). While there seems to be agreement that analogies are an integral part of human expertise there is difficulty with this approach in finding usefully linked domains or analogies, or relevant mappings between those and real situations (Buchanan, 1989). Of late, neural nets have become a popular concept. While they are reasonably successful in domains where there is little knowledge required e.g. perceptual tasks, there is no definitive way to understand why some cases yield success and others failure (ibid).

2.4 Criteria for selecting experts

A key factor in the success of an ES project is selecting the right expert(s) on whom to base the elicitation of expertise. An expert is defined in The Collins English Dictionary as

"a person who has extensive skill or knowledge in a particular field."

While there is no doubt that considerable amounts of expertise are held by individuals within an organisation, some researchers contend that this is an unnecessarily limited view. Klein (1992), suggests that a more logical approach is not so much one of "who is an expert" but rather "where is the expertise in a given organization." He goes on to note that criteria such as years of service or formal qualifications are less valid indicators than peer group assessment of an individual's credibility in a domain. On occasions, the expert may be supplied by the organisation without the knowledge engineer having any input into selection. Not all experts will be suitable for involvement in KA activities, and it falls to the knowledge engineer to screen potential participants. Some inherent dilemmas have been identified which can be likened to the old adage about 'painters houses never getting painted'. Parsaye & Chignell (1988) note that often, the impetus for building an ES is created by a shortage of experts in the domain. As a result, experts are frequently inaccessible, frustrating attempts to achieve continuity in KA activities. Some essential characteristics are:

- a. possession of appropriate specialised knowledge which sets them apart from novice performers.
- b. the ability to recognize the boundaries of their knowledge.
- c. provision of solutions in a timely fashion.
- d. not attempting to solve problems where their specialised knowledge does not apply.
- e. available, articulate and committed.

2.5 Use of multiple experts

2.5.1 General

The subtle nature of expertise means that problems may be solved by a variety of experts through the use of different techniques. Boose (1984) notes that two experts may use similar or different methods to arrive at similar or different solutions, all of which may be acceptable. It is notable that the literature has shown a distinct shift toward the use of techniques involving multiple experts in the last two years (ref).

2.5.2 Need for multiple experts

The need for the use of multiple experts is argued by Boose and Gaines (1989) because they believe that it is unlikely that one person will possess all the knowledge required in a complex domain. Ford and Adams-Webber (1992) describe two reasons for using multiple domain experts. First, because of the problem of narrow expertise, also known as the fragility problem, where an expert system fails because its knowledge is too limited. Second, that no one expert may be capable of adequately solving the problem. These two issues are referred to respectively as issues of length and breadth.

2.5.3 Examples of the use of multiple experts

Shaw & Woodward (1988) conducted research using *KSSO* as the KA tool to measure the degree of consistency of multiple experts between entities and attributes, and over time. They found that in the first study using academics, experts were at no time fixed in their views of the topic and were accepting of other ways of considering the problem. This group used the terminology in much the same way over time. The second study involved practising professionals in the field who were generally much more decisive over procedure but demonstrated some difference in the use of terminology. This highlights the effect on KA of the type of expert and the nature of the domain. Having used abnormal sensor readings to drive a forward-chaining rule based ES, Payne and McArthur (1990) described

the integration of a second expert's knowledge using backward chaining rules to work from a hypothesis to a cause. Comparing the models of multiple experts is, in itself, a useful elicitation technique. This concept has been carried further into validation strategies for ES. The Expertise Transfer System (*ETS*) (Boose, 1984) is an attempt to allow the user to make selections among multiple expert models.

Some domains or particular circumstances make it desirable for different knowledge bases to be maintained separately at all times. These are known as blackboard systems. Blackboard systems are those where multiple knowledge sources are held in partitions, only interacting via a scheduler. Solutions are built up on the blackboard which is a very complex working memory for the system (Jackson, 1990). This type of system is ideal where different elements of expertise need to be integrated case by case, where it is desirable to allow the user to select the knowledge source, or where consensus cannot be reached between multiple experts on some issues. The earliest attempt at implementing this type of system was in a speech-understanding program called *Hearsay* (Erman, Hayes-Roth, Lesser & Reddy, 1980) which attempted to integrate knowledge about signal processing, speech waveform mapping to natural language and search for meaning.

2.5.4 Problems of multiple expertise sources

The inherent diversity in expert knowledge presents a problem for KA as each expert will likely employ or articulate a unique model of their knowledge. Experts may also be in disagreement over issues such as vocabulary and definitions and so the knowledge engineer is faced with the problem of trying to build a consensus model without losing knowledge specific to individual experts. The fundamental psycho-social principles which make KA such a challenge grow combinatorially as more than one expert joins the team and interactions become more complex. It has been noted that a KA plan will not be as prescriptive for multiple experts as for

one (Hickman, Killin, Land, Mulhall, Porter, & Taylor, 1989). This adds more time to the process than just the additional KEL requires. Brule and Blount (1989) note that while much is known about human learning and elicitation techniques, few researchers have attempted to bring the two areas of study under one framework. They cite the relationship between the expert(s) and knowledge engineer as an essential area for study.

2.6 Conclusion

There is a growing body of literature covering all aspects of the field of knowledge acquisition. The rate of contribution has accelerated markedly in the last 10 years and particularly so in the last five. Some major difficulties exist in the field of knowledge acquisition which are revealed by a review of the literature. First, the field is wide-ranging and so research effort is diminished by not focusing on one angle or issue. Second, there is no general agreement on vocabulary, approaches or standards for the discipline. This makes it difficult to relate different results and experiments and is made more so by the security necessary because of the commercial potential of expert systems. Despite this, some generic standards and agreements are being made and the KADS program is an example of this.

Because of the proliferation of meanings, it is important now to establish some working definitions on which to base this study. This study employs knowledge elicitation techniques and the definition to be used is that of Cordingley (1989) shown earlier. Discourse analysis is defined as the analysis of transcripts in order to identify key concepts, the relationships between them and the attributes of both.

An introduction to the domain of interest, sensory evaluation, including significant definitions is given in chapter 3.

Chapter 3

Sensory Evaluation Panels

3 Sensory Evaluation Panels

3.1 Introduction

Humans have always relied on their senses to provide them with information about the environment. Over time, human desire has become more sophisticated and so the senses are no longer used just for issues of survival, but to choose between items of equal safety, in order to select that which provides the greater satisfaction. While the process has existed informally since the beginning of mankind, sensory evaluation as a formal science is a relatively recent discipline. Although some studies were carried out as early as the 13th century (Jennings, 1977), it wasn't until after World War 2 that the field really developed, spurred on by problems with ration acceptability amongst soldiers.

3.2 Sensory evaluation defined

The field of sensory evaluation has been defined by the US Institute of Food Technologists' Sensory Evaluation Division as:

"a scientific discipline used to evoke, measure, analyze and interpret sensations as they are perceived by the senses of sight, smell, taste, touch and hearing" (Gatchalian, 1981).

In practical terms, sensory evaluation is carried out for commercial reasons to determine quality or acceptability of a product. It is normally done through the use of the human senses, although advances are being made in the development of other tests that measure factors objectively (Larmond, 1977). Although a basic form of sensory evaluation is made by the scientist or food technologist at the time that a new food product is developed, most assessment is carried out by panels. Larmond (1977) groups panels into three types; highly trained experts, laboratory panels, and large consumer panels. The expert is most often used to assess quality, as in the case of the master winemaker. Large consumer panels,

perhaps up to a hundred or more strong, are used to determine consumer reaction to a product. The research which is the subject of this thesis, focuses on trained laboratory panels, which are suitable for control purposes or identification of changes.

3.3 Selecting panellists

In order to achieve the highest possible level of objectivity, potential panellists are screened before being committed to the training process. It is difficult, however, to set rigid guidelines and depending on the pool of potential panellists, relative performance is sometimes the most suitable basis for selection. (Giradot, Peryam and Shapiro, 1952). The ability to sense important attributes of the target product has usually been the focus of panellist selection. However there are many other important criteria including the person's individual attributes. These include communication skills, ability to function within a group, motivation and reliability. Traditionally, panellists have been drawn from the staff of the research establishment but this is now giving way to a more commercially oriented approach, where panellists are recruited 'off-the-street' from all walks of life. The establishment which is the subject of this research conducts this style of recruitment.

3.4 Training panellists

The individual members of the panel are the means of analysis in sensory evaluation, and so it is important that they are trained not only to be able to identify product attributes but also to act as controls on confounding external factors. Before a panel can be committed to commercial project work, a period of training is necessary. This training is normally conducted by the group leader. Even trained panels need regular revision training in order to keep them current and also to reinforce standards. The training of laboratory panels is conducted during scheduled sessions and focuses on developing skills in attribute recognition for the particular product that the panel is to be used on. As such, no two panel training processes are the same and there is no general agreement on how long the training process should take.

3.5 Panel leaders

Each panel is headed by a panel leader. This person is a highly skilled expert with a relevant tertiary background such as food science or psychology. Panel leaders are appointed at the outset to raise, train and manage a panel for a specific project or product type. The autonomous nature of this task and the individual expertise the leader brings to it give rise to variations in both process and outcome for the panel. While research unit managers may lay down guidelines for selection and training, the panel leader, at the end of the day, has almost total control of the panel.

3.6 Statistical aspects of panels

Measurement is the process of assigning numerical values to empirical observations (Cascio, 1990). This is the only valid means by which the consistency and reliability of the collected data can be established, and it is on this data that subsequent commercial decisions are based. Consequently, panel leaders must design activities which produce parameter scores that lend themselves to statistical analysis (Gatchalian, 1981). Statistical results are used as part of the selection and training process as a means of identifying outliers and providing feedback to panellists. Computer software has been engineered that enhances this process, enabling panellists to input responses directly and be provided with results straight after a session. Where this facility does not exist, panel leaders must analyze the data and provide feedback.

3.7 Suitability as a domain

The domain selected for any knowledge engineering process will have a profound effect on the ease and success of elicitation. Several potential domains in the field of business, agriculture and veterinary science were considered before finally selecting NZ DRI to make an approach to. The domain was selected on two grounds. First, its nature is inherently suitable for knowledge engineering as it deals with high levels of individual expertise, ambiguity and choice. Second, the organisation and staff of the target organisation were all very motivated to make

the project successful and this came to be a key finding in predicting the success of such an undertaking. It is unusual to find several experts in the same field all working together in the same organisation. After initial discussions, it was clear that individual procedures varied and this also indicated the usefulness of a KA project. Further details of the domain and the individual experts are provided in chapter 4.

Chapter 4

The Knowledge Acquisition Method

4 Methodology

4.1 General

The nature of this study was exploratory. It was designed to examine issues that will affect the acquisition of knowledge from multiple experts in the same domain. The approach was based on the KADS model employing discourse analysis techniques to assist in the development of the domain layer of the knowledge base. The domain selected was that of panellist training for sensory evaluation, using three experts who are all employed by the New Zealand Dairy Research Institute. All three have either current or recent experience with training sensory evaluation panellists. The author acted as the knowledge engineer throughout the study.

4.2 Detailed methodology

The knowledge acquisition methodology employed in this research followed the modelling approach proposed in KADS (Wielinga et al, 1991) with a project goal of developing the domain layer constructs of concepts and relations between concepts. Interviews were transcribed in full and the transcripts analyzed using discourse analysis techniques (Kuipers & Kassirer, 1987). Individual concept lists were combined into one set during a consensus meeting of all the experts with the knowledge engineer and this was used as the basis for developing a domain model and schema. A detailed description of all aspects of the methodology is provided in the remainder of this chapter.

4.2.1 The domain

Sensory evaluation is the process of using trained human subjects to assess such factors as flavours, textures and aromas in food products. This field is more fully described in Chapter 3. It was considered a suitable area for expert systems development because of the extent of the ambiguity inherent in flavours, the level of human decision making involved and the relative scarcity of human experts. There is little definitive literature on

exactly how to train a sensory panel and so experts, with backgrounds ranging from food science to psychology, are used to train panels for a particular food group. Having assessed the field as suitable for research, a proposal was put to the manager of the Sensory Evaluation Unit which was accepted subject to constraints on commercial confidence. A copy of the proposal is shown at Appendix A to this thesis.

Within the domain of sensory evaluation, the greatest area of ambiguity lies in the area of training of sensory panellists. This is, by and large, a one to many activity between a panel leader, such as the experts in this research, and up to twenty members of the public. In particular, attention was focused on the activities of the training cycle that panel leaders undertake to establish whether a panel is ready to commence commercial project work or not.

4.2.2 The experts

Three experts were used in this research. The demographical and biographical word pictures for each are shown in Appendix G. All three are employed by the Sensory Evaluation Unit (SEU) of the New Zealand Dairy Research Institute (DRI) based in Palmerston North. Expert 1 was the SEU manager and was the most experienced of the three. While she was not training panellists during this research, she had considerable experience in doing so in the past. She was also responsible for the supervision of all panel training and often assisted other experts or stood in, in cases of their absence. Expert 2 was the next most experienced trainer. In her post as technologist, she had trained panels for a wide variety of food groups and was involved with a chocolate panel during this research. Expert 3, also a technologist, was the least experienced of the three but also had worked within a wide variety of food groups. Her primary responsibility during the period of this research was in training panellists in evaluating baked products. It is pertinent to note that panellists are more highly motivated to sample pleasant tasting foods such

as baked products rather than basic items such as butter, but it is inversely proportional to the skill needed on the part of the expert who must deal with very complex compositions in the former.

4.2.3 The interviews

The primary method of knowledge elicitation was via semi-structured interviews. This type of interview, where the line of questioning pursues a similar but non-rigid path toward the same objective is more fully described in chapter 2.

Each expert was interviewed on two separate occasions for no more than one hour. The interviews were recorded on dictaphone and transcribed into hard copy. Transcription was deemed necessary because of the volume and complexity of the verbal data which would have made analysis directly from tapes extremely difficult. Transcripts were initially handed back to the interviewees for proof-reading and for the insertion of words inaudible on the original tape. Each expert was requested not to show their transcripts to other participants in the research at this stage, due to the potential to skew the result of the consensus meeting, discussed in paragraph 4.2.6. Contamination of data also refers in this situation to the potential for the knowledge engineer to become too knowledgeable about the field from one expert and therefore alter his or her approach to subsequent experts. The interviews were conducted in sequence with each expert having an equal amount of interview time before the process moved on a stage. (Hickman, Killin, Land, Mulhall, Porter and Taylor, 1989). Proofed transcripts were collected in and checked to ensure that the same ground had been covered for each expert. This was found to be the case.

4.2.4 Initial analysis

The interview tapes were transcribed and are shown in full as Appendix B to this thesis. The transcripts of the interviews were initially analyzed based on the discourse analysis method expounded by Kuipers & Kassirer (1987).

This method was further enhanced to involve the experts in the analysis process. This was designed to increase their feeling of involvement and consequently their commitment to the success of the project. First, each expert was briefed individually by the knowledge engineer on the analysis procedure. The experts were coached in identifying excerpts in the narrative where they appeared to be focusing on factual explanation rather than expressing opinion about their own mental processes. After going through two to three pages of the original transcript and satisfying himself that each expert understood the process of identifying excerpts, the full transcripts were left with respective experts for them to complete the identification of such excerpts. These edited transcripts were collected in the next day and cross-checked. This checking involved the knowledge engineer reading through the edited versions to ensure that the remaining content reflected factual material only, and that potentially important items had not been removed. This was found to be the case. The excerpts used in this research are shown at Appendix C. An example of the excerpt process is shown in Tables 1 and 2 on pages 28 and 30.

KE	Have the group leaders been given feedback about their variation in the same way that the panellists do?
Expert 1	No. The impression amongst the experts, I think, is that they're all the same. There's no doubt in my mind, however, that this is not the case as there are definite differences in assertiveness and how hard each one pushes a panel
KE	I see. What is the general feeling amongst the group leaders toward this knowledge acquisition project?
Expert 1	Overall, it's positive. I think they're a little unsure about what's expected of them but all see this as part of the overall quality programme
KE	Good. It sounds to me like the training process is open-ended, time wise?
Expert 1	Yes. It used to be 3 months with Establishment 1 staff using laymen who are not familiar with the field take longer, as do larger numbers. There are standard deviations for each product based on historical data but leaders aren't using them so my question is what are they basing their decisions on? You see the length of panel training ranges at present from 4 to 6 months, depending on how motivated and keen they are. It also depends on how good the leaders standards are and how pushy they get

Table 1: Example of full transcript page prior to analysis

KEI	
2	Have the group leaders been given feedback about their variation in the
3	same way that the panellists do?
4	... No... there are definite differences in assertiveness and how hard each
5	one pushes a panel
6	... There are standard deviations for each product based on historical data
7	but leaders aren't using them so my question is what are they basing their
8	decisions on? You see the length of panel training ranges at present from
9	4 to 6 months, depending on how motivated and keen they are. It also
10	depends on how good the leaders standards are and how pushy they get

Table 2: Example of factual excerpt following initial analysis

4.2.5 Further analysis

The second stage of the analysis involved the knowledge engineer working with the reduced excerpts and breaking them into short lines that correspond more or less to a meaningful phrase or action. This greatly simplifies further analysis by revealing, in most cases, only one concept or relation per line (Kuipers & Kassirer, 1987). The excerpts reduced to phrase lines are shown at appendix D to the thesis. An example of the state of the transcript after this exercise is shown in Table 3 on page 30. The next stage of the analysis took place in two parts. The first was to identify the concepts in the domain that the expert was referring to. Concepts are the basic building blocks of a knowledge base and represent prototypical items or their associated properties (Jackson, 1990). Examples of the concepts derived from this research are 'selection' or 'threshold testing'.

KEI

- 2 Have the group leaders been given feedback about their variation in the
3 same way that the panellists do?
- 4 No... there are definite differences in assertiveness
5 and how hard each one pushes a panel.
- 6 There are standard deviations for each product based on historical data
7 but leaders aren't using them
8 so my question is what are they basing their decisions on?
- 9 You see the length of panel training ranges at present from 4 to 6 months,
10 depending on how motivated and keen they are.
- 11 It also depends on how good the leader's standards are
12 and how pushy they get

Table 3: Example of excerpt broken into meaningful phrases

A general reading of the transcripts indicated the context and major foci of the experts' knowledge. Five group headings; stages of training, training activities, training standards, selection activities and other concepts/properties were deduced and the knowledge engineer set about identifying concepts. The group headings provided the framework for the conduct of the overall process. Concepts and properties were noted as they were identified, with the line numbers being recorded for every occurrence. An example of a partial concept list following this stage is shown in Table 4 on page 31. Each expert's transcript was analyzed separately in the first instance and then comparisons were drawn. A comparison chart was constructed to identify areas of similar or different content. Where the knowledge engineer believed that different words meant the same thing, the concepts were placed on the same line but enclosed in square brackets for later verification with the experts. This was an essential step in the analysis, for two different reasons. First, as explained in chapter 2, many experts have difficulty articulating their knowledge. During elicitation

therefore, it may be that the meaning of the data is different from the

<u>Stages of Training</u>
Selection (L2)
Training (L12,166)
Decision point (L30/31, 35, 114, 122, 307-315, 352)
Project work (L130, 162-163, 187, 369)

<u>Training Activities</u>
Threshold testing (L3, 88)
Triangle tests (L7, 88)
Product description (L13, 18, 206, 210, 345-346, 359, 360)
Words to numbers (L23)
Feedback (L29, 39, 58, 59, 250-253)
Group dynamics (L95,102, 127, 134, 178, 261)
Training session (L100, 104-106, 166, 366, 368)
Analysis of performance (L110-113, 115)
Half-blind tests (L155-160, 172)

Table 4: Example of partial object list derived from excerpts

actual usual meaning attributed to a word. Comparisons enable these differences to be addressed. Second, consensus was an important outcome for this project. The end-users had made it clear that they wanted only one knowledge base and so the comparison chart served as the base document upon which the consensus meeting would focus. An example of the combined concepts chart is shown as Table 5 on page 32.

COMBINED CONCEPTS			
NO	EXPERT 1	EXPERT 2	EXPERT 3
TRAINING ACTIVITIES			
6.	Triangle tests	-	-
7.	Product description	[Word generation]	[Word generation]
8.	Words to numbers	Words to numbers	-
9.	Feedback	Feedback	Feedback

Table 5: Example of combined concepts chart for multiple experts

4.2.6 Development of consensus

Following the analysis of individual concept vocabularies and development of the combined concepts chart, all three experts were brought together to compare their approaches in a meeting facilitated by the researcher. The achievement of consensus, as mentioned earlier, was a vital outcome for the SEU as they desired one knowledge base on which to base their training and quality management in the future. To this end, all activities throughout the elicitation process were oriented toward creating a team approach, consensus, and commitment to a successful outcome. This included such things as social visits by the knowledge engineer to the unit for morning teas, taking an interest in the out of work activities of all staff at the unit, and the provision of consultancy services which were not directly related to the project itself. An example of this type of activity was the conduct of a personality type survey and explanatory workshop. The consensus meeting is therefore only part of the overall function of building commitment but is the step in that function which, by its nature is the most fully documented.

The transcript of this 1 hour 30 minute meeting is shown as Appendix F to this thesis. Through discussion between the experts regarding the processes employed and the exact meaning of certain phrases, a single set of concepts was agreed on. The second part of the analysis was then conducted by the knowledge engineer. This was to identify the various hierarchies and relationships between the concepts. The agreed concept vocabulary was used as the foundation for development into model form based on the domain layer of the KADS methodology (Wielinga et al, 1991).

The KE process was concluded at this point and an analysis of the findings from the KA process was conducted. This is described fully in chapter 6.

4.3 Conclusion

The methodology employed in this project has brought together facets of the work of several researchers. It is focused around the KADS assumption that knowledge elicitation is a modelling activity and subsequent findings will show this to be an accurate assumption with respect to this domain. Knowledge from the social sciences led the knowledge engineer to pursue an approach which was designed to enhance the group identity of the three experts and build a firm commitment to the success of the project. This was felt to be the best way to overcome many of the problems in reaching consensus inherent in the use of multiple experts and more fully discussed in chapter 2. This commitment building programme was extremely successful and the results are discussed in subsequent chapters.

Chapter 5

Analysis of the Transcripts

5 Analysis

5.1 Interviews

Any examination of the analysis of verbal data should start with the actual collection at source. This is necessary because it is at the interview stage that the knowledge engineer carries out the initial analysis. Through gauging the expert's demeanour and responses, it is possible to filter out unusable material via skilful questioning. Resistance to the project, if identified early, can be dealt with during the initial data collection stages by reassuring, supporting and sometimes confronting the expert.

In this particular project, there was obvious keenness to participate. However, there were difficulties identified in the initial interview with each expert which warranted special attention from the knowledge engineer. These ranged from talkativeness, requiring questioning to be brought back to the point often, as well as shy or vague responses which required coaxing to draw out key points.

5.2 Transcript analysis

As mentioned earlier, the sheer volume of the verbal data made analysis from tapes impractical. A Massey University Research Fund Grant of \$500 enabled a secretarial service to be employed to produce transcripts of the nine individual interviews. This hard copy was then used as the basis for discourse analysis. The analysis of the transcripts was conducted in accordance with the techniques employed by Kuipers and Kassirer (1987) and Wielinga et al (1991). This involves two main steps:

- a. Identify excerpts where the expert appears to be concentrating on or explaining knowledge (rather than expressing opinion about own mental processes).

- b. For each excerpt: identify the concepts, properties and relations between them (as distinct from the wording referring to them) and identify the causal relations that are described.

5.2.1 Phase 'A' analysis

Phase A analysis was conducted with the involvement of the three experts. This was not only in order to build commitment and understanding toward the project but also substantially reduced the level of manual labour for the knowledge engineer, who was working without a support team. This phase of the analysis was initiated via a teaching seminar at the SEU. Each expert was coached in the method of identifying factual segments in their own transcript and then left to mark blocks of text where they thought they were identifying knowledge. The following day, marked transcripts were collected and checked by the knowledge engineer.

5.2.2 Phase 'B' analysis

This phase involved the knowledge engineer only. The reduced excerpts were broken down into short lines that correspond roughly to meaningful phrases in the explanation. The phrases were initially read for general meaning. It was identified that the experts were talking, in general terms about stages of training, training activities, training standards, selection activities and other objects which for the time being were identified simply as that. It is pertinent to note that the general headings are flexible and in this case they evolved with the analysis. They are primarily to provide focus for the knowledge engineer to group like concepts and relations. The excerpts were analyzed line by line under the general headings mentioned to identify individual items and those, along with the transcript line numbers of their occurrences were recorded. The outcomes from the process form the basis of this chapter.

5.3 Expert 1

The excerpts for expert 1 are shown at Appendix B. An initial analysis revealed

the following concepts and properties:

Stages of Training

Selection (L2)

Training (L12,166)

Decision point (L30/31, 35, 114, 122, 307-315, 352)

Project work (L130, 162-163, 187, 369)

Training Activities

Threshold testing (L3, 88)

Triangle tests (L7, 88)

Product description (L13, 18, 206, 210, 345-346, 359, 360)

Words to numbers (L23)

Feedback (L29, 39, 58, 59, 250-253)

Group dynamics (L95,102, 127, 134, 178, 261)

Training session (L100, 104-106, 166, 366, 368)

Analysis of performance (L110-113, 115)

Half-blind tests (L155-160, 172)

Scale development (L203, 204, 241, 243, 244, 261, 268, 273, 277-281, 282-285, 290, 362, 365)

Product evaluation (L246-247, 291-295, 316-318, 320, 328, 344, 348)

Training Standards

Taste discrimination (L4, 5, 8, 32, 271-272)

Time (L9, 17, 21, 26, 44, 49, 50, 53, 146-150, 161, 169, 172, 177, 184, 307, 350, 352, 368)

Product attributes (L14, 215, 219, 227-228, 326, 329, 330, 334, 340-343, 347, 360, 365)

End of scale values (L19, 20, 264)

Variation (L27, 34, 35, 164, 192)

Standard deviation (L28, 41, 153, 202, 209, 212-214, 218, 219, 222, 224-225, 249, 286-287,

308)

Mean (L151, 248)

F sets (L305)

Selection Activities

Interviewing (L62-65, 72, 81, 141, 143)

Data collection (L67-70, 73-80)

Decision to hire (L83-93)

Screening tests (L88, 302-304)

Other Concepts and Properties

Training environment (L6, 11, 41, 56, 57, 67-70, 157, 159, 177, 189, 201, 242, 243, 244)

Population size (L9, 10, 91)

Group leader (L16, 19, 25, 35, 42, 46, 99, 109, 120-121, 122, 370, 372-376)

Pay (L51, 52)

5.4 Expert 2

The excerpts for expert 2 are shown at appendix B. An initial analysis revealed the following concepts and properties:

Stages of Training

Selection (L2-31, 294)

Decision Point (L27-28, 100, 101, 496, 565,580)

Training (L32)

Project Work (L386, 548-551, 581-582)

Training Activities

Word generation (L45, 59, 481-482)

Panel protocols (L54-57)

Reference sample development (L62, 276, 483)
Scale development (L67, 69, 76, 494, 556-557)
Words to numbers (L74)
Feedback (L81, 226, 256, 369, 406)
Training session (L284, 395, 403, 415-416, 462)
Group co-ordination (L335, 510-511)
Session facilitation (L343, 348)
Researching (L345)
Analysis of performance (L354, 360, 420-421, 456, 513, 534, 544)
Project samples (L377)

Training Standards

Time (L4, 44, 142-144, 147-148, 149-150, 177-178, 181-183, 283, 562)
Reproducible (L79, 119)
Consistency (L80, 251-265, 275)
Fully trained (L118-123)
Standard deviations (L120, 130-131, 134-135, 138-139, 423, 426, 440, 536, 563, 566)
Range (L368, 563, 566)
Reference samples (L396, 441)
Mean (L536, 563)

Selection Activities

Advertise (L2)
Screening sessions (L3, 5, 211, 311-320)
Identify basic tastes (L7)
Threshold testing (L8-10)
Triangle tests (L12-15)
Descriptive work (L16-19)
Fun activities (L20-21)
Personality assessment (L22)

Interview (L25, 329)
Data analysis (L330)
Removal (L496)

Other Concepts and Properties

Population size (L29, 109-117)
Panel/ Product attributes (L33-41, 95, 158-162, 163, 252, 258, 265, 270, 408-409, 442, 484, 502)
Commercial pressure (L123, 127, 242-246, 285, 435)
Group leader (L124, 128, 167-176, 289-291, 303-309, 360, 361-363, 444, 513, 525-529)
Pay (L198-202, 204, 213)
Panel expense (L218-222)
Sensitivity (L277-278)
Compusense (L554, 562)

5.5 Expert 3

The excerpts for expert 3 are shown at appendix B. An initial analysis revealed the following concepts and properties:

Stages of Training

Screening (L4, 52-54)
Training (L15, 195, 238-241)
Project work (L186)

Training Activities

Group dynamics (L17)
Word generation (L20, 22, 109, 116-117)
Scale development (L29-30, 38-39, 46, 129, 140, 212-215)
Reference sample development (L31, 198)

Booths (L48, 261)
Panel protocols (L101-105, 123-128)
Feedback (L111-112, 153, 169, 300-301)
Sample production (L118-119)
Group interaction (L137)
Analysis of performance (L152, 178, 188, 201)
Training blocks (L228-229)

Training Standards

Consistency (L49, 203, 210-211, 272)
Competency (L154)
Standard deviation (L156-159, 172, 193-194, 211)
Mean (L156, 193, 196)
Time (L166, 168, 206, 266, 268-269, 320)
ANOVA (L192)
Test exposures (L288)

Selection Activities

Advertise (L2)
Screening tasks (L4-5, 57)
Identify basic tastes (L6, 67)
Difference testing (L7, 61, 69)
Personality work (L13, L81-83)
Group dynamics (L14, 17)
Questionnaire (L50)
Time (L58, 59, 87-88)
Threshold testing (L63)
Triangle testing (L71)
Description tests (L74)
Interview (L76)

Session (L79)

Other Concepts and Properties

Population size (L2, L92-93, 97)

Panel/Product attributes (L9, 26, 178)

Group leader (L41-42), 180-181, 248-253)

Pay (L175)

5.6 Consensus vocabulary

Following the identification of individual concepts and properties, a meeting with all three experts, facilitated by the KE developed a shared vocabulary which was agreed on by consensus. This meeting was based on the 'combined concepts chart' shown at Appendix E. The shared vocabulary was developed into a rough hierarchy and sequence during the meeting and is shown below.

SHARED VOCABULARY IN SEQUENCE FOLLOWING CONSENSUS MEETING OF EXPERTS

Stages of Training (Top level)

1. Planning
2. Selection
3. Training
4. Decision point

Planning Activities (Middle level)

There is only one activity placed in this category, having been transferred here from line 20 of the initial combined concepts chart. This does not indicate that there are no more,

only that the top level stage of planning was developed late in the KA process and the focus of the study had already been placed on the decision as to when a panel was considered trained. The relocated concept is included here for completeness.

(1a) Project samples

Selection Activities (Middle level)

5. Advertise
6. Demographic data collection
7. Pre-screening questionnaire
8. Screening tests
9. Threshold tests
10. Discriminatory tests
11. Descriptive work
12. Group dynamics evaluation
13. Individual panellist evaluation
14. Performance evaluation
15. Decision to hire

Included in Selection Activities in no set order

- (16) Fun activities
- (17) Selection session

Training Activities (Middle level)

18. Panel protocols
19. Administration
20. Session facilitation
21. Warm-up activities
22. Descriptor generation

23. Descriptor development
24. Training sample development
 - 24a. Discriminatory tests
25. Scale development
26. Reference sample development
 - 26a. Discriminatory tests

Included in training Activities in no set order

- (27) Analysis of group performance
- (28) Analysis of individual performance
- (29) Feedback
- (30) Training session

Performance Measures (Bottom level: order of no consequence)

31. Taste discrimination
32. Time
33. Consistency
34. Blind references
35. Standard deviation
36. Mean
37. Range
38. Graphical representation

Other Considerations (Order of no consequence)

39. Product complexity
40. Training environment
41. Population size
42. Group leader
43. Pay

44. Panel make-up
45. Commercial pressure
46. Panel expense
47. Data collection method

5.7 Development of domain layer model

The aim of the modelling phase of this research was to engineer the elicited knowledge, using a modelling approach. There are numerous ways of dealing with modelling and a review of these is covered in Chapter 2. As a complete model would have taken considerably more acquisition and development, it was deemed beyond the scope of this study, which was primarily focussed on knowledge elicitation issues. To achieve a degree of closure on the knowledge already acquired, it was decided to progress to a domain model and review the issues from that point.

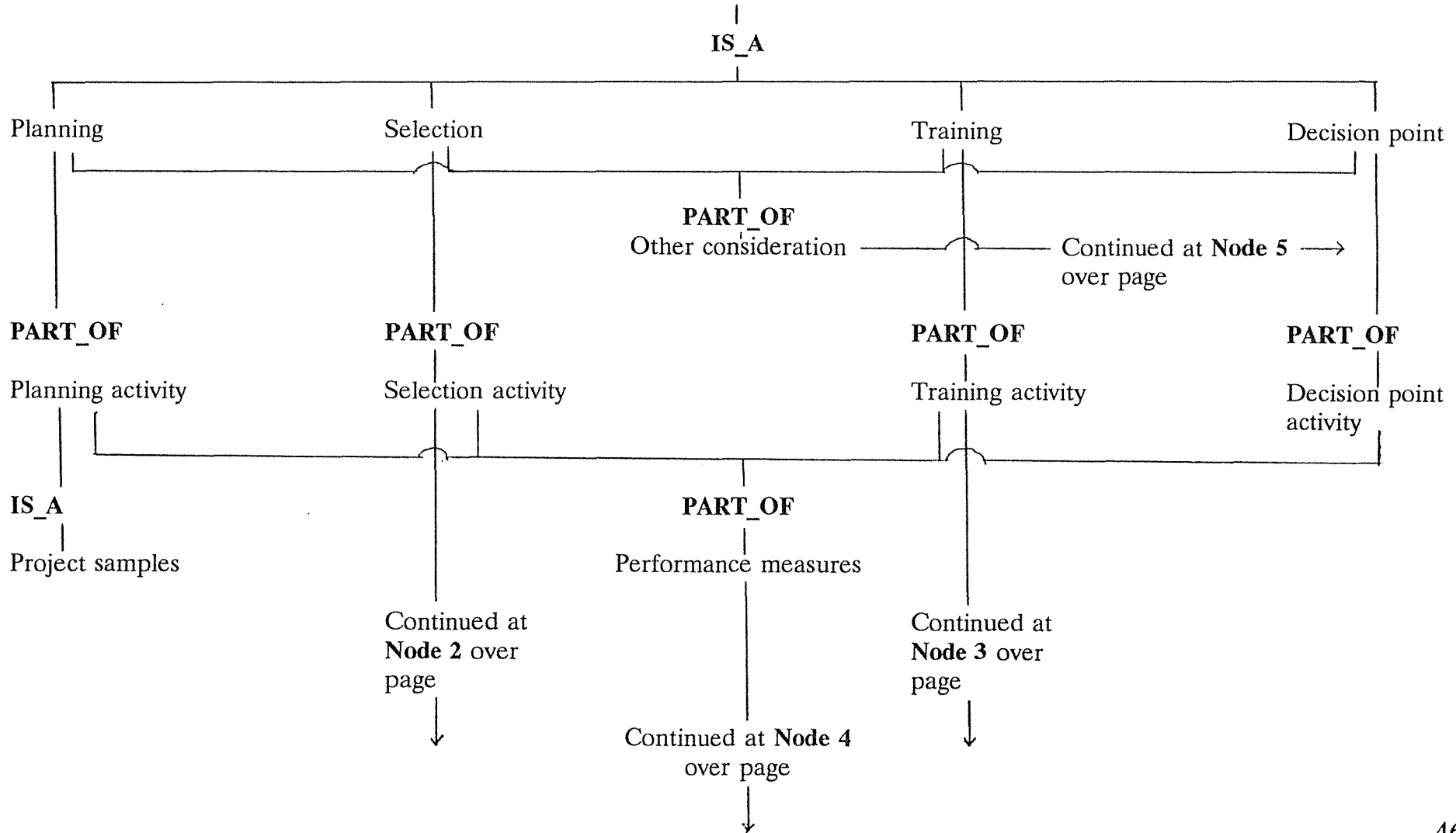
The purpose of the domain model is viewed in slightly different ways by different researchers. Kuipers and Kassirer (1987) describe it as "having the purpose of making explicit, information that is logically necessary to answer questions about the domain, but may not have been stated in the explanation." The KADS "four layer model" methodology views the domain model as "the static knowledge of the domain: domain concepts, relations and complex structures, such as models of processes or devices" (Schreiber, Breuker, Bredeweg and Wielinga, 1988). The separation into four layers reflects the different ways that knowledge can be viewed, especially in terms of content versus controlling knowledge. In the KADS domain layer there are four ontological primitives which are concepts, properties, relations between concepts and property expressions (Wielinga, Schreiber and Breuker, 1991). These primitives enable the specification of a domain schema.

The shared vocabulary of concepts and properties was used as a basis for the development of a domain model. This was based on the KADS (Wielinga et al, 1991) approach as described above. The domain schema diagram shows the hierarchical relationships between concepts in a more pictorial form. It is shown

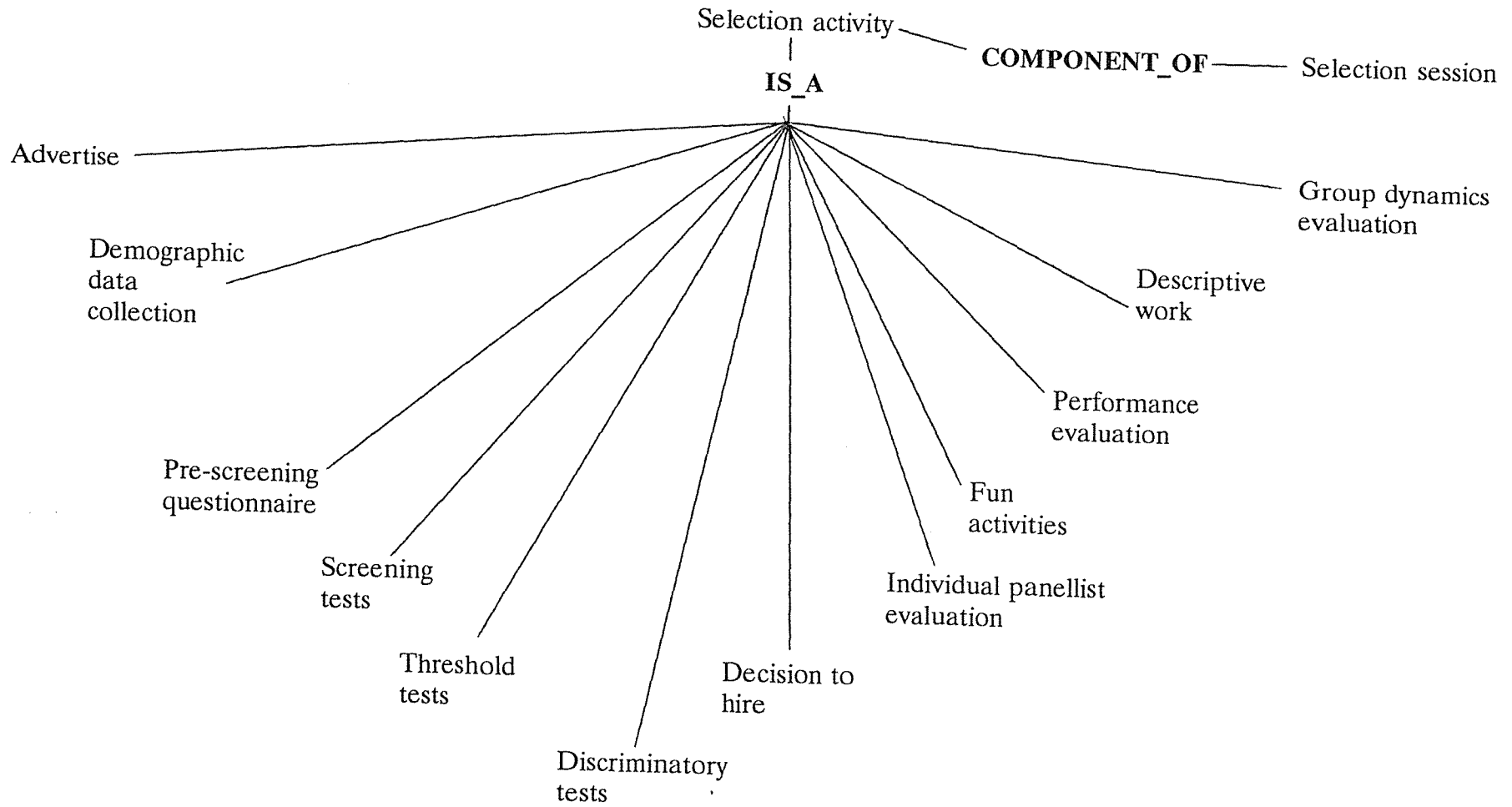
on page 46.

DOMAIN MODEL DIAGRAM

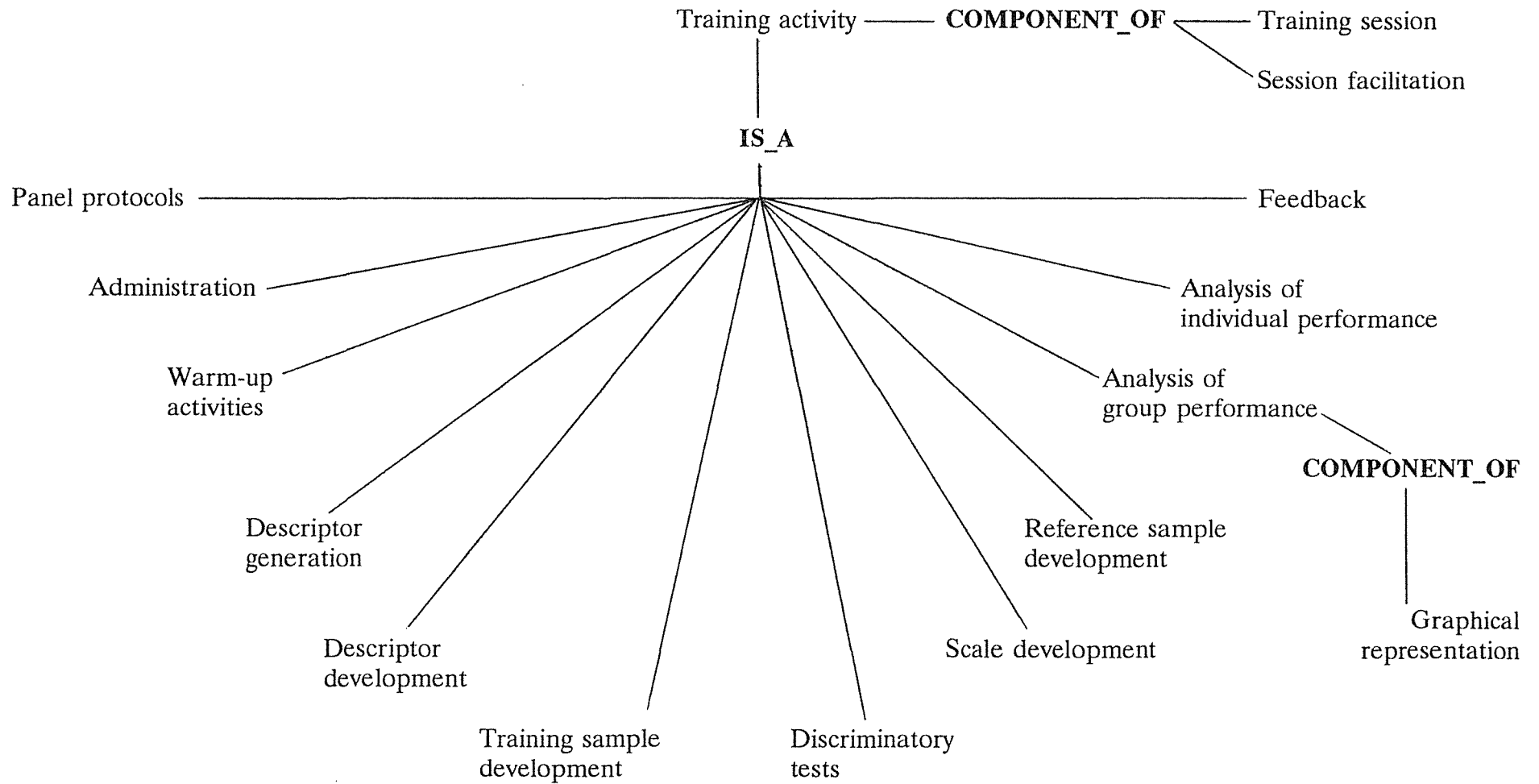
STAGES OF TRAINING



NODE 2



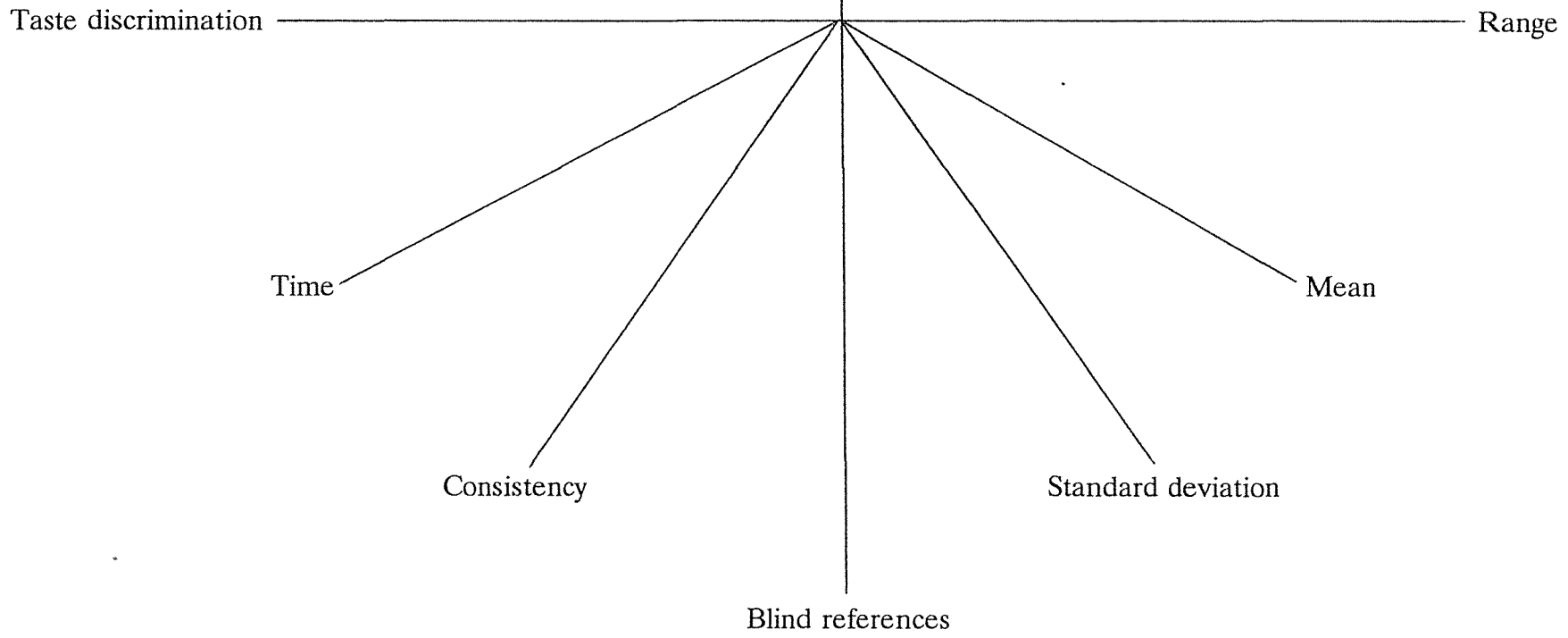
NODE 3



NODE 4

Performance measure

IS_A



NODE 5

Other consideration

IS_A

Product complexity

Data collection

Training environment

Panel expense

Population size

Commercial pressure

Group leader

Panel make-up

Pay

No	Name	Description
	CONCEPTS	
	Stages of training	The stages that a group leader goes through in training a panel
	Planning	Group leaders must plan the training of the panel to which they are assigned
	Selection	Members of the panel are specially selected according to certain criteria
	Training	Members of the panel are trained by their group leader to pre-determined performance standards
	Decision point	There is a set of activities and a certain point in time where the group leader must decide whether the panel is ready for project work
	Activities	There are activities which take place at different stages within panel training
	Project samples	The samples needed for different stages of training
	Advertise	Ads are placed to attract potential panel members
	Demographic data collection	The applicant basic facts are recorded on a database
	Pre-screening questionnaire	Further data collected at the time the selection process starts
	Screening tests	A range of basic tests
	Threshold tests	
	Discriminatory tests	Test of the applicant's ability to discern between tastes
	Descriptive work	The applicant learns to describe product attributes
	Group dynamics evaluation	An assessment of how the group fits together
	Individual panellist evaluation	An assessment of the personal suitability of each applicant

No	Name	Description
	Performance evaluation	A more objective measure of the applicant's ability to taste differences
	Decision to hire	Selecting applicants to proceed to panel training
	Fun activities	Other activities to relieve boredom during selection
	Selection session	The basic unit of time during which a selection activity is conducted
	Panel protocols	The procedural aspects of belonging to a panel
	Administration	Developing training records, pay data etc
	Session facilitation	The group leader facilitates the group's processes
	Warm up activities	Getting the group members comfortable with each other
	Descriptor generation	Generating words to reflect samples in use
	Descriptor development	Developing words into an agreed vocabulary
	Training sample development	Agreeing on standard reference samples for certain descriptors
	Performance measures	The criteria used to establish performance in certain stages of training
	Taste discrimination	The ability to differentiate between basic tastes
	Time	Time taken for an event
	Consistency	Degree to which results are reproduced over time
	Blind references	Ability to discern basic samples
	Standard deviation	A measure of group variance around the mean
	Mean	The average result of the group
	Range	The spread of scores of the group
	Graphical representation	The method by which performance is fed back to the group
	Other considerations	Important issues which are not activities

No	Name	Description
	Product complexity	The number of product attributes and the ease with which they can be separated
	Training environment	Venues such as round table room or individual booths
	Population size	The number of people involved in a particular activity
	Group leader	The personal characteristics of the group leader
	Pay	Reimbursement for attending activities
	Panel make-up	The attributes of the panel as a whole
	Commercial pressure	The pressure to produce commercial sensory evaluation results
	Panel expense	The cost of convening and training a panel
	Data collection method	The method of collecting results
	RELATIONS BETWEEN CONCEPTS	
	Planning IS_A Stage of Training	Part of the hierarchy of activities in panel training
	Selection IS_A Stage of Training	Part of the hierarchy of activities in panel training
	Training IS_A Stage of Training	Part of the hierarchy of activities in panel training
	Decision point IS_A Stage of Training	Part of the hierarchy of activities in panel training
	Activity PART_OF activity	Sub type generic activity of range of activities
	Planning activity PART_OF planning	The range of activities which take place during planning
	Selection activity PART_OF selection	The range of activities which take place during selection
	Training activity PART_OF training	The range of activities which take place during training
	Decision point activity PART_OF decision point	The range of activities which take place during decision point

No	Name	Description
	Project samples IS_A planning activity	Group leader plans project samples for training
	Advertise IS_A selection activity	One of activities conducted during selection
	Demographic data collection IS_A selection activity	One of activities conducted during selection
	Pre-screening questionnaire IS_A selection activity	One of activities conducted during selection
	Screening tests IS_A selection activity	One of activities conducted during selection
	Threshold tests IS_A selection activity	One of activities conducted during selection
	Discriminatory tests IS_A selection activity	One of activities conducted during selection
	Descriptive work IS_A selection activity	One of activities conducted during selection
	Group dynamics evaluation IS_A selection activity	One of activities conducted during selection
	Individual panellist evaluation IS_A selection activity	One of activities conducted during selection
	Performance evaluation IS_A selection activity	One of activities conducted during selection
	Decision to hire IS_A selection activity	One of activities conducted during selection
	Fun activities IS_A selection activity	One of activities conducted during selection
	Selection session COMPONENT_OF selection activity	Basic time period for activities conducted during selection
	Panel protocols IS_A training activity	One of activities conducted during training
	Administration IS_A training activity	One of activities conducted during training

No	Name	Description
	Session facilitation COMPONENT_OF training activity	Ongoing facilitation by group leader
	Warm-up activities IS_A training activity	One of activities conducted during training
	Descriptor generation IS_A training activity	One of activities conducted during training
	Descriptor development IS_A training activity	One of activities conducted during training
	Training sample development IS_A training activity	One of activities conducted during training
	Discriminatory tests IS_A training activity	One of activities conducted during training
	Scale development IS_A training activity	One of activities conducted during training
	Reference sample development IS_A training activity	One of activities conducted during training
	Discriminatory tests IS_A training activity	One of activities conducted during training
	Analysis of group performance IS_A training activity	Group leader activity during training
	Analysis of individual performance IS_A training activity	Group leader activity during training
	Feedback IS_A training activity	Group leader activity during training
	Training session COMPONENT_OF training activity	Basic time period for activities conducted during training
	Performance measures PART_OF planning activities	Group leader pre-determines measurement criteria

No	Name	Description
	Performance measures PART_OF selection activities	Group leader pre-determines measurement criteria
	Performance measures PART_OF training activities	Group leader pre-determines measurement criteria
	Performance measures PART_OF decision point activities	Group leader pre-determines measurement criteria
	Performance measure IS_A performance measure	Sub group type
	Taste discrimination IS_A performance measure	One of performance measures used
	Time IS_A performance measure	One of performance measures used
	Consistency IS_A performance measure	One of performance measures used
	Blind references IS_A performance measure	One of performance measures used
	Standard deviation IS_A performance measure	One of performance measures used
	Mean IS_A performance measure	One of performance measures used
	Range IS_A performance measure	One of performance measures used
	Graphical representation COMPONENT_OF feedback	Means of providing performance feedback to group
	Other considerations PART_OF planning	Part of planning stage
	Other considerations PART_OF selection	Part of selection stage
	Other considerations PART_OF training	Part of training stage
	Other considerations PART_OF decision point	Part of decision point

No	Name	Description
	Product complexity IS_A other consideration	Other consideration in each stage of training
	Training environment IS_A other consideration	Other consideration in each stage of training
	Population size IS_A other consideration	Other consideration in each stage of training
	Group leader IS_A other consideration	Other consideration in each stage of training
	Pay IS_A other consideration	Other consideration in each stage of training
	Panel make-up IS_A other consideration	Other consideration in each stage of training
	Commercial pressure IS_A other consideration	Other consideration in each stage of training
	Panel expense IS_A other consideration	Other consideration in each stage of training
	Data collection IS_A other consideration	Other consideration in each stage of training

5.8 Conclusion

Analysis of verbal data is an extremely laborious and time-consuming activity. In this research, it represented a 'bottleneck' far greater than was anticipated. This labour intensive aspect was exacerbated by the lack of guidance in the literature as to the amount of time it takes to transcribe and analyze an interview of one hour. Had this data been available, planning could have taken place to arrange more concurrent activity. Second, a single knowledge engineer is pushed to deal with three experts on the same project and a support team would have facilitated productivity by carrying out administrative tasks. The introduction of other staff brings its own penalty, however, in terms of management time and discontinuity of work.

Involving the experts in the analysis provide both labour saving and built commitment to the project. This is believed to be a vital factor in the

development of a single knowledge base from multiple experts. A full examination of the findings from this research are presented in chapter 6.

Chapter 6

Discussion of the Findings

6 Discussion

6.1 Introduction

The findings from this research are discussed in chronological order. This does not allocate appropriate emphasis to key findings and to cater for this, a summary of key findings is provided below. The general discussion is restricted to matters relevant to the research goal posed. It is acknowledged that secondary analysis may well identify matters arising which pertain to areas other than the acquisition of knowledge from multiple experts in the same domain.

6.2 Key findings

The significant findings from this study, together with the reference to their full discussion are:

- a. *Organisational commitment is vital to successful knowledge engineering projects.* Paragraph 6.3 discusses the advantages to be gained from working within an organisation that demonstrates its commitment to the success of knowledge engineering projects.
- b. *Motivated experts facilitate the elicitation of knowledge.* Paragraph 6.4 identifies the benefits for the knowledge engineer in dealing with experts who are personally and professionally motivated.
- c. *Administrative support is necessary for a project involving multiple experts.* Paragraph 6.9 highlights the labour intensive nature of dealing with verbal data and examines the pros and cons of administrative assistance.
- d. *Every opportunity must be taken to educate and involve the experts in the knowledge engineering process.* Several paragraphs in this chapter discuss the involvement of the experts in various stages of the process. This builds commitment to the overall success of the project.
- e. *Bringing the experts together in a consensus meeting is of immense*

value to the overall success of the project. Various options are open to the KE to synthesize the knowledge of the many into a single model. Paragraph 6.10 examines the options and outcomes for the consensus stage.

6.3 Organizational commitment

The domains considered as potential research areas all satisfied the generally agreed requirements for successful knowledge engineering in that they contained elements where expertise was scarce, and there was ambiguity or a lack of agreement of the exact method for achieving a task. The key factor which led toward the approach to DRI was organisational commitment. Many other potential organisations appeared disinterested or suspicious of the concept whereas, the Sensory Evaluation Unit (SEU) staff at DRI were a young, motivated group who were keen to find new and better ways of understanding how they did their work. The manager of the SEU saw potential for the research to contribute to the DRI Total Quality programme and helped to gain approval for this project from higher management. Organisational commitment is deemed vital to the success of knowledge engineering. To this end, specific commitment building exercises were incorporated into the research design and these are discussed fully in paragraph 6.6.

6.4 Identification of experts

Having gained entry to an organisation, the literature suggests several methods by which the experts who are to be involved may be identified. In this research project the selection was not problematic. There were three experts available, one of whom was the SEU manager. All were keen to be involved in the project, albeit that at the outset they had little understanding of the scope but, it is felt were carried along by the enthusiasm of the manager. A significant difference between experts at the start was the level of experience. The more experienced experts appeared to have a more conscious grasp of the processes and activities they were involved in. More junior staff needed careful questioning to bring out the same points. While selection of experts is an important factor for successful

knowledge engineering, their ongoing availability is equally important, particularly in projects of long duration. The provision of motivated experts is a facilitating factor in the elicitation of knowledge and undoubtedly these experts were easier to work with than if they had been forced to comply or indifferent to the project. Due to difficulties in finding times when both knowledge engineer and expert were both available the completion date was moved back on several occasions. This had an impact on the availability of the experts. The unit manager was posted internally within DRI and, as she was a key part of the expert team, the closing stages of the elicitation and modelling process had to be compressed to achieve closure. It is recommended that in projects of medium to long duration, all required times are block-booked. This facilitates project management and is a superior and more systematic approach rather than conducting knowledge elicitation sessions on an 'as available' basis.

6.5 Research design

Because of the nature of this study and the relative lack of literature using multiple experts, the design of this research was based on aspects of the work of several different researchers. This proved to be a strength, as it provided a dynamic approach, drawing on the best features of various designs. No single piece of reported research provided a prescriptive approach to the process of building a single knowledge base from several experts. There were many publications dealing with elicitation techniques, and the analysis and modelling of the data. A few identified issues related to the involvement of multiple experts in the same domain and still fewer on the integration of their knowledge. Interviewing has been well documented over the years as has the analysis of verbatim transcripts as data. When working with multiple experts, it was felt that structured interviews were too rigid, given that different experts will take different cognitive paths to the achievement of the same outcome. As such, semi-structured interviews were found to be the most suitable means of interviewing multiple experts. The procedures followed by Ericsson and Simon (1984) and Kuipers and Kassirer (1987) proved to be very suitable for analysis of the transcripts in this project. What has developed from this study by combining supporting techniques

with the KADS methodology is a possible method for eliciting and modelling multiple expertise into a single consensus knowledge base.

6.6 Organisational and individual commitment

The level of commitment of both the organisation and individual has been commented on earlier. Because this factor showed up early as being important to the process, specific activities were developed to enhance the feeling of involvement of the participants. These included a general discussion of the field of knowledge engineering, regular updates on the research which were normally conducted informally at morning or afternoon tea in the SEU offices, and personality type testing. The latter activity employed the Personal Styles Inventory (Hogan & Champagne, 1980) which employs a Jungian (1921) typology with sixteen possible outcomes. Although its applicability to some fields of human resource management is questionable, it does give individuals another means of understanding their own unique attributes. In the team context, those who better understand their own differences are often more tolerant of the differences of others, thereby enhancing the performance of the group. It is pertinent to note that commitment building exercises were not limited to the three experts but involved all staff of the unit. This was designed to offset the feeling of exclusion for support staff and to gain commitment to the final outcome as a valuable resource for the unit. It further motivated the three experts by enhancing their status once support staff understood the scope and purpose of the project.

6.7 Processing of interview data

All interviews were recorded on a dictaphone and transcribed into full written accounts. This was an extremely labour intensive activity. A limited resource was available for this research and so a secretarial service was hired to do some of the transcription while the knowledge engineer did the remainder. This was a poor but unavoidable use of time. Each one hour interview tape would take approximately eight hours to transcribe. The use of a proper transcription machine, with headset and foot operated tape buttons reduced this time to about 6 hours per 1 hour of tape. Nonetheless, the key finding seems to be that effective

support teams are vital to cope with the sheer volume of data generated in the elicitation phase. An administrative assistant would have been an asset to this project and would have reduced total project time of two years, by two to three months. There is a cost to the involvement of other support staff when implementing this finding. All staff will consume a certain percentage of time in management. Also, when more than one person is involved with analysis of verbal data, the potential for inconsistencies is increased, and this may cause difficulties in achieving a consensus knowledge base from multiple experts. After the initial transcription was complete, the scripts were handed back to the experts, under a non-disclosure arrangement, for checking. Sections that were inaudible on the tape were inserted and errors of interpretation were corrected. This was not only beneficial from the point of view of accuracy of data, but was also another opportunity to involve the expert, further building their commitment to the project.

6.8 Reduction of transcripts

The full transcripts were reduced to only those excerpts where the expert is revealing actual knowledge, rather than opinion or general discussion. Knowledge engineer comments were only left in where they were considered vital to the context of the answer provided. Whilst the transcripts could have been reduced by the knowledge engineer, it was felt that this was another opportunity to build commitment via the involvement of the experts. Each expert was individually coached in how to differentiate between factual knowledge and opinion. To ensure that they understood it, they were each asked to reduce the first page or two while the knowledge engineer was present. Once satisfied with the standard, they were then left to complete the task, without reference to each other's work. Reduced transcripts were collected back in the next day. To ensure accuracy, the knowledge engineer checked the excerpts to ensure they accurately reflected the desired content. The excerpts were then broken into single statements which roughly corresponded to meaningful phrases. This was a purely mechanical activity to facilitate discourse analysis and this could easily be done by an assistant or by a software application. It is clear from re-reading the transcripts that this is an

essential step in simplifying the data prior to attempting to identify concepts and relations. Neither phase of the reduction should be eliminated from future work.

6.9 Discourse analysis of individual transcripts

From single line statements the knowledge engineer identified words or phrases that seemed to indicate concepts or relationships. This is an extremely subjective process and it was found that it had to be done several times with increasing degrees of refinement. A general reading was necessary to establish the broad categories of terminology. Once an initial vocabulary had been identified, it was then necessary to go back over the list in order to find duplications of concepts on more than one group. It was found that three 'passes' per expert were needed to establish a conceptual vocabulary representative of the transcript excerpts without duplicating or omitting any activities of importance. When working toward an outcome of a single knowledge base, it was felt that it was better, in the initial stages, to leave in any individual experts' statement which was in doubt, in case that this had meaning for the other experts at a later stage. This proved to be the case. During the consensus meeting, the experts queried each other on the exact meanings of words when they all had not mentioned it in the interviews. They were often surprised with their own omissions and agreed on what the group definition would be.

6.10 Achievement of consensus

Ultimately, the three individual expert vocabularies needed to be reduced to one and it was anticipated that, as the literature suggests, this would present some difficulties, due to the individualistic nature of expert knowledge. This was not found to be the case. There were several options open to the knowledge engineer regarding the achievement of consensus. Enforcement by the organization was discounted early as it would have undermined commitment to the end-use product. Leaving the experts alone to work through the differences would have taken much longer and individual personalities might still have prevailed. Ultimately, consensus through knowledge engineer intervention was chosen. The consensus approach was very successful as shown by the fact that all difficulties

relating to terminology and procedures were able to be resolved amicably during the consensus meeting. It is believed that this is due, at least in part, to the team spirit of the SEU, as well as the specific activities conducted over the preceding eighteen months by the knowledge engineer, designed to enhance commitment to the project.

6.11 Benefits to the experts

All three experts reported high levels of satisfaction over their involvement in the project. Specific benefits that were identified included:

- a. more conscious comprehension of their own knowledge,
- b. an increased understanding of the processes employed by their colleagues, and
- c. the opportunity to question and review individual and organizational procedures which were, in some cases, found to be used more out of tradition than utility.

6.12 Benefits to the organisation

In discussions with the three experts and with other staff of the SEU, it was evident that the project had resulted in tangible benefits for the organization. The organizational benefits discussed included:

- a. a well documented hierarchy of activities for the training of panellists which could be used for the briefing of future panel trainers,
- b. a coherent vocabulary upon which to base the ongoing development of quality assurance programmes, and
- c. enhanced unit spirit through involvement in a project which is seen to be a front runner in the sensory evaluation field.

6.13 Conclusions

The research goal was to identify issues relating to the acquisition of knowledge from multiple experts. Despite the literature suggesting this to be problematic, the only major difficulty was one of scale, in processing the volume of data elicited. The knowledge acquisition procedures themselves work well but there is clearly

much more work to be done in understanding human behaviour. The impact of the social sciences is discussed more fully in chapter 7.

Chapter 7

Conclusions

7 Conclusions

7.1 A modelling approach

The KADS modelling approach was selected from amongst the range of possible knowledge acquisition methods due to the fact that it is free from any form of knowledge representation and emphasises a construction rather than mining approach. This was an appropriate choice given that it was clear from the outset that the project would not proceed to the construction of a knowledge base. The organization still required a usable outcome and the KADS modelling approach provided several stages at which the research could be halted without detriment to the utility of the knowledge already elicited.

7.1.1 The value of manual elicitation

The decision to employ manual rather than automated elicitation techniques proved justified. The major findings from this research are based around human interaction and these would not have been revealed had the experts interacted with a computer during knowledge acquisition instead of the knowledge engineer. Too little is known at this stage about the personal processes of human experts and this suggests that, while software acquisition tools may be of assistance in reducing the laborious nature of knowledge engineering, they should not be used to totally replace the knowledge engineer.

7.2 Discourse analysis

Discourse analysis has been conducted in different forms for many years, although it has gained general acceptance as a means of using verbal reports as data over the last ten years with the work of researchers such as Ericsson & Simon (1984). While some later researchers have employed the theories of predicate calculus in representing concepts and relationships in the analysis stage, it was found that this was at odds with the KADS modelling approach. Consequently, concepts and

relationships were derived directly from the text, as described by Kuipers & Kassirer (1987). The techniques of discourse analysis are now well established and this stage of the project presented no real problems for the researcher. It is considered to be a suitable analysis technique for use with modelling approaches and is equally suitable for use with multiple experts.

7.3 Issues relating to multiple experts

A review of the literature revealed several different ways in which the knowledge of multiple experts can be elicited and employed. It has been found through this research that the method employed will be, in whole or part, a function of the ultimate use intended for the knowledge. If the knowledge is from different areas, the user needs to know the specific source or expert consensus is impossible, then a blackboard approach to representation will be preceded by individual elicitation programmes. In the case of this research, the user required a single knowledge base which reflected the combined expertise of three individuals. From the outset, therefore, it was clear that the elicitation plan had to be oriented toward ultimately achieving consensus. This is believed to be an important factor in the planning considerations for acquisition of knowledge from multiple experts in the same domain.

7.3.1 Contamination potential

The potential for contamination due to the experts talking to each other was high but it was pleasing to note from the surprise at individual comments in the consensus meeting that the experts did not discuss their individual results with each other. Another potential problem which did not manifest itself was interference at the consensus stage due to the relative status of the experts. With one of the three being the manager of the others, it was possible that views on concepts and relationships would be imposed but this did not occur. It is an issue that the knowledge engineer must be aware of and measures should be planned to make allowance for this possibility in similar work.

7.4 Organizational issues

The factors of greatest significance revealed in this study revolve around the theories which seek to understand what energises and sustains human behaviour. The literature indicated many potential difficulties in working with multiple experts. It was therefore surprising to experience few difficulties in eliciting knowledge and achieving consensus. It is felt that the commitment of all levels of management to the project, the innovative culture and the high personal levels of motivation of the experts were all key facets of the success of this project. Future projects of this nature would do well to emphasise these issues as essential design features, both as predictors of success and also elements to be enhanced during the project.

7.5 Redefining the role of the knowledge engineer

As a logical corollary of the findings of this research, it is felt that the traditional view of the KE's role may require re-definition. The traditional view has held the KE as a person who is primarily skilled in utilising knowledge elicitation techniques, modelling that knowledge and implementing it in the desired form. This research has shown that, to a large extent, well managed experts can do a lot of those tasks themselves. The counterpoint to this is that, no matter how skilled the knowledge engineer may be in the actual techniques of elicitation, an inability to deal on the human and organisational level with the participants in the project may well limit the likelihood of success. The training of KE in the future might need to emphasise these skills, accepting the position that the actual elicitation techniques are well proven.

7.6 Directions for future research

This work was focused from the outset on the knowledge acquisition processes required to model the expertise of more than one expert in the same domain. It has been found that the two major techniques used, KADS modelling and discourse analysis are well proven and very suitable for knowledge acquisition tasks. What has become evident from the research is the inseparability of expert systems in general and knowledge acquisition in particular, from the disciplines of

social science. It is felt that an understanding of the acquisition of knowledge from multiple experts in the same domain could be further enhanced by a re-examination of the field from a social science perspective. Topics which are believed to be relevant for future study include:

- a. personality type of experts and its relationship to the conscious articulation of knowledge,
- b. personality type of the knowledge engineer and the match between it and that of the experts,
- c. the relationship between the skill and experience levels of the knowledge engineer and experts in communication activities and the success of the project, and
- d. the relationship between the organizational culture, the level of organizational commitment and the ultimate success of the project.

7.7 Closing summary

Though laborious, this research is considered a success on two counts. First, it has delivered to the user organization the benefits which it promised. It is felt that providing deliverables is vital to the future of knowledge engineering. If the field cannot win the commercial support necessary to move out of the laboratories then its growth will be limited. Second, this two and a half year project has achieved its research goal. As is often the nature of exploratory research, the findings were not those that were anticipated at the outset. The researcher set out expecting to discover difficulties with elicitation and modelling techniques when they were applied to multiple experts but this was not the case. The major issues to arise were, in fact, a function of the success of the project, highlighting the need for greater understanding by computer scientists of the theories which seek to explain the behaviour of their human subjects. Like most exploratory research, it is felt that further work is required before a definitive conclusion can be made from the findings of this work.

References

- Akkermans, H., Van Harmelen, F., Schreiber, G. & Wielinga, B. (1992). A formalisation of knowledge-level models for knowledge acquisition. *International Journal of Intelligent Systems*. Vol 8, No. 2 Feb.
- Barr, A. & Feigenbaum, E. (1981). *The handbook of artificial intelligence*, Vol 1, California: Heuristech Press.
- Barr, A. & Feigenbaum, E. (1982). *The handbook of artificial intelligence*, Vol 2, California: Heuristech Press.
- Belkin, N.J. & Brooks, H.M. (1987). Knowledge elicitation using discourse analysis. *International Journal of Man-Machine Studies*. **27**, 127-144.
- Bennet, J.S. (1985). ROGET: A Knowledge-based system for acquiring the conceptual structure of a diagnostic expert system. *Journal of Automated Reasoning*, **1**, 49-74.
- Boose, J.H. (1984). Rapid acquisition and combination of knowledge from multiple experts in the same domain, in *Artificial intelligence applications; the engineering of knowledge-based systems*. C. R. Weisbin (ed). Massachusetts: IEEE Computer Society Press.
- Boose, J.H. (1985). A knowledge acquisition program for expert systems based on personal construct psychology. *International Journal of Man-Machine Studies*. **23**, 495-525.
- Boose, J.H. & Bradshaw, J.M. (1987). Expertise transfer and complex problems using AQUINAS as a knowledge acquisition workbench for knowledge-based systems. *International Journal of Man-Machine Studies*, **26**, 3-28.

- Boose, J.H. & Gaines, B. (1989). Knowledge acquisition for knowledge-based systems: Notes on the state-of-the-art, in *Knowledge acquisition: selected research and commentary*. S. Marcus (ed). Massachusetts: Kluwer.
- Brule, J.F. & Blount, A. (1989). *Knowledge Acquisition*. New York: McGraw-Hill.
- Buchanan, B.G. (1989). Can machine learning offer anything to expert systems? In *Knowledge acquisition: Selected research and commentary*. S. Marcus (Ed.). Boston: Kluwer.
- Buis, M., Hamer, J., Hosking, J.G. & Mugridge, W.B. (1986). An expert advisory system for a fire safety code, in *Applications of expert systems*, Vol 1. Quinlan J.R.(Ed), Wokingham: Turing Institute Press.
- Burton, A.M., Shadbolt, N.R., Rugg, G., and Hedgecock, A.P. (1988). Knowledge elicitation techniques in classification domains. In Y. Kodratoff (Ed.), *ECAI-88: Proceedings of the Eight European Conference on AI* (pp 85-90). London: Pitman.
- Bylander, T. & Chandrasekaran, B. (1987). Generic tasks for knowledge-based reasoning: the "right" level of abstraction for knowledge acquisition. *International Journal of Man-Machine Studies*, **26**, 231-243.
- Cascio, W.F. (1991). *Applied psychology in personnel management*. 4th ed. New Jersey: Prentice-Hall.
- Clancey, W.J. (1985). Heuristic classification. *Artificial Intelligence*, **27**, 289-350.
- Cleaves, J.G. (1987). Cognitive biases and corrective techniques: proposals for improving elicitation procedures for knowledge-based systems. *International Journal of Man-Machine Studies*. **27**, 155-166.
- Collins English Dictionary (1986). Glasgow: William Collins and Sons.

- Cooke, N.M. & McDonald, J.E. (1987). The application of psychological scaling techniques to knowledge elicitation for knowledge-based systems. *International Journal of Man-Machine Studies*, **26**, 533-550.
- Cordingly, E.S. (1989). Knowledge elicitation techniques for knowledge-based systems. In Diaper, D. (1989). *Knowledge elicitation principles, techniques and applications*. Chichester: Ellis Horwood.
- Davis, R. (1977). Interactive transfer of expertise: Acquisition of new inference rules. *International Journal of CAI* **5**, 321-328.
- Duda, R.O. & Gaschnig, J.G. (1981). Knowledge-based expert systems come of age. *Byte*, **6**, 238-281.
- Ericsson, K.A. & Simon, H.A. (1984). *Protocol analysis: Verbal reports as data*. Massachusetts: MIT Press.
- Erman, L., Hayes-Roth, F., Lesser, V. & Reddy, D. (1980). The HEARSAY-II speech understanding system: integrating knowledge to resolve uncertainty. *Computing Surveys*, **12**(2), 213-253.
- Eshelman, L., Ehret, D., McDermott, J. & Tan, M. (1987). MOLE: A tenacious knowledge-acquisition tool. *International Journal of Man-Machine Studies*, **26**, 41-54.
- Ford, K.M. and Adams-Webber, J.R. (1992). Knowledge acquisition and constructivist epistemology. In *The psychology of expertise*, Hoffman, R.R. Ed. New York: Springer-Verlag.
- Gaines, B. & Linster, M. (1990). Development of second generation knowledge acquisition systems. In *Current trends in knowledge acquisition*. B. Wielinga et al (Eds.). Amsterdam: IOS Press.

- Gammack, J.G. (1987). Different techniques and different aspects on declarative knowledge. In A. Kidd (Ed), *Knowledge acquisition for expert systems* (pp. 137-164). New York: Plenum.
- Garg-Janardan, C. & Salvendy, G. (1988). A structured knowledge elicitation methodology for building expert systems. *International Journal of Man-Machine Studies*. **29**, 377-406.
- Gatchalian, M.M. (1981). *Sensory evaluation methods with statistical analysis*. College of Home Economics, University of the Philippines.
- Giradot, N.F., Peryam, D.R. and Shapiro, R. (1952). Selection of sensory testing panels. *Food Technology*, 6 (4) pp. 140-143.
- Hart, A. (1992). *Knowledge Acquisition for expert systems*. 2nd Ed. New York; McGraw-Hill.
- Hayes-Roth, F., Waterman, D.A. & Lenat, D. (1983). *Building expert systems*, Reading Massachusetts: Addison-Wesley.
- Hogan, R.C. and Champagne, D.W. (1980). Personal Style Inventory. In Pfeiffer and Jones, (1980). Eds, *The 1980 Annual handbook for group facilitators*. University Associates.
- Hickman, F.R., Killin, J.L., Land, L. Mulhall, T., Porter, D. and Taylor, R.M. (1989). *Analysis for knowledge-based systems: A practical guide to the KADS methodology*. London: Ellis Horwood.
- Jackson, P. (1990). *Introduction to expert systems*, 2nd Ed, Wokingham: Addison-Wesley.

- Jennings, W.G. (1977). Objective measurements of flavor quality; General approaches, problems, pitfalls and accomplishments. *ACS Symposium series, No. 51*. Scanlan, R.A. Ed.
- Jung, C.G. (1921). Psychological Types. In Read, H. et al (Eds), *The collected works of CG Jung*. (1953). London: Routledge, Regan & Paul.
- Karback, W., Linster, M. & Voss, A. (1990). Models of problem-solving: One label - one idea? In Wielinga, B. et al eds. (1990). *Current trends in knowledge acquisition*. Amsterdam: IOS Press.
- Kelly, G.A. (1955). *The psychology of personal constructs*. New York: Norton.
- Klein, G.A. (1992). Using knowledge engineering to preserve corporate memory. *In The psychology of expertise*, Hoffman, R.R. Ed. New York: Springer-Verlag.
- Klinker, G.J., Bentolila, S., Genetet, S., Grimes, M. & McDermott, J. (1987). KNACK-Report-driven knowledge acquisition. *International Journal of Man-Machine Studies*. **26**, 65-79.
- Kodratoff, Y. & Manago, M. (1987). Generalization and noise. *International Journal of Man-Machine Studies*. **27**, 181-204.
- Kuipers, B. & Kassirer, J.P. (1987). Knowledge acquisition by analysis of verbatim protocols. *in Knowledge acquisition for expert systems. A practical handbook*. (1987). Kidd, A.L. (Ed) New York: Plenum.
- LaFrance, M. (1987). The knowledge acquisition grid: a method for training knowledge engineers. *International Journal of Man-Machine Studies*, **26**, 245-255.
- Larmond, E. (1977). *Laboratory methods for sensory evaluation of food*. Research Branch, Canada Department of Agriculture. Publication 1637.

- Lefkowitz, L.S. & Lesser, V.R. (1988). Knowledge acquisition as knowledge assimilation. *International Journal of Man-Machine Studies*. **29**, 215-226.
- Marcus, S. (1988). *Automating knowledge acquisition for expert systems*. Boston: Kluwer.
- Minsky, M. (1975). A framework for representing knowledge. In Winston, P.H. ed, (1975). *The psychology of computer vision*. New York: McGraw-Hill.
- Musen, M.A., Fagan, L.M., Combs, D.M. & Shortliffe, E.H. (1987). Use of a domain model to drive an interactive knowledge-editing tool. *International Journal of Man-Machine Studies*. **26**, 105-121.
- Musen, M.A. (1989). *Automated generation of model-based knowledge-acquisition tools*. London: Pitman.
- Newell, A. & Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.
- Nunez, M. (1990). Decision tree induction using domain knowledge. In *Current trends in knowledge acquisition*. B. Wielinga et al (Eds.). Amsterdam: IOS Press.
- Parsaye, K. & Chignell, M. (1988) *Expert systems for experts*. New York: Wiley.
- Payne, E.C. and McArthur, R.C. (1990). *Developing expert systems: A knowledge engineer's handbook for rules and objects*. New York: Wiley.
- Pettit, E.J. (1987). Analysis of the performance of a genetic algorithm-based system for message classification in noisy environments. *International Journal of Man-Machine Studies*. **27**, 205-220.
- Quinlan, J.R., Compton, P.J., Horn, K.A. & Lazarus, L. (1986). Inductive knowledge acquisition: A case study. In J.R. Quinlan (Ed.), *Applications of expert systems*.

Maidenhead: Academic Press.

- Ruqian, L. & Cungen, C. (1990). Towards knowledge acquisition from domain books. In *Current trends in knowledge acquisition*. B. Wielinga et al (Eds.). Amsterdam: IOS Press.
- Schmidt, G. & Schmalhofer, F. (1990). Case oriented knowledge acquisition from texts. In *Current trends in knowledge acquisition*. B. Wielinga et al (Eds.). Amsterdam: IOS Press.
- Schreiber, G., Breuker, J., Bredeweg B., & Wielinga, B. (1988). Modelling in KBS development. *Proceedings of the second European knowledge acquisition workshop EKAW '88*. Bonn.
- Shadbolt, N. & Wielinga, B. (1990). Knowledge based knowledge acquisition: the next generation of support tools. In *Current trends in knowledge acquisition*. B. Wielinga et al (Eds.). Amsterdam: IOS Press.
- Shalin, V.L., Wisniewski, E.J., Levi, K.R. & Scott, P.D. (1988). A formal analysis of machine learning systems for knowledge acquisition. *International Journal of Man-Machine Studies*. **29**, 429-446.
- Shaw, M.L.G. & Gaines, B.R. (1987). KITTEN: Knowledge initiation and transfer tools for experts and novices. *International Journal of Man-Machine Studies*. **27**, 251-280.
- Shaw, M.L.G. & Woodward, J.B. (1988). Validation in a knowledge support system: construing and consistency with multiple experts. *International Journal of Man-Machine Studies*. **29**, 329-350.
- Shortliffe, E.H., Scott, A.C., Bischoff, M.B., van Melle, W., and Jacobs, C.D. (1981). ONCOCIN: An expert system for oncology protocol management. *Proceedings of the 7th International Joint Conference on Artificial Intelligence*. pp 876 - 881.

Sommerville, I. (1989). *Software engineering*. 3rd ed. Wokingham: Addison-Wesley.

Wielinga, B.J., Schreiber, A. Th., & Breuker, J.A. (1991). KADS: A modelling approach to knowledge engineering. *KADS-II/T1.1/PP/UvA/008/1.0 Public*. University of Amsterdam.

Wilkins, D.C., Clancey, W.J. & Buchanan, B.G. (1987). Knowledge base refinement by monitoring abstract control knowledge. *International Journal of Man-Machine Studies*. **27**, 281-293.

APPENDICES TO
THE ACQUISITION OF KNOWLEDGE FROM MULTIPLE EXPERTS
IN THE DOMAIN OF
SENSORY EVALUATION PANEL TRAINING

Contents

A.	Research Proposal	2
B.	Full transcripts	
	Full transcript Expert 1	11
	Full transcript Expert 2	36
	Full transcript Expert 3	70
C.	Excerpts	
	Expert 1 excerpts	100
	Expert 2 excerpts	107
	Expert 3 excerpts	118
D.	Excerpts by meaningful phrase	
	Expert 1: excerpts by meaningful phrase	125
	Expert 2: excerpts by meaningful phrase	143
	Expert 3: excerpts by meaningful phrase	169
E.	Combined concepts chart	185
F.	Transcript of consensus meeting	192
G.	Expert biographical data	229

Appendix A

Proposal for Research

PROPOSAL-IN-PRINCIPLE

FOR THE CONDUCT OF
RESEARCH INTO EXPERTISE HELD BY
MULTIPLE EXPERTS IN THE SAME DOMAIN

To be carried out by

Simon Ewing-Jarvie
Assistant Lecturer in Human Resource Management
Faculty of Business Studies
Massey University

at

The Sensory Evaluation Unit
of the
New Zealand Dairy Research Institute



**MASSEY
UNIVERSITY**

Private Bag
Palmerston North
New Zealand
Telephone 0-6-356 9099
Facsimile 0-6-350 5608

**FACULTY OF
BUSINESS STUDIES**

DEPARTMENT OF
HUMAN
RESOURCE
MANAGEMENT

15 April 1992

Ms Sharon Wards
Section Head
Sensory Evaluation Unit
NZ DRI
Private Bag
PALMERSTON NORTH

Dear Sharon

Thank you for the opportunity to further discuss the potential for my research to be carried out at the Sensory Evaluation Unit.

As I mentioned when we spoke, the research I am about to undertake forms the basis of a thesis which will result in the completion of a Master of Philosophy degree in computer science. My particular focus is in the area of knowledge acquisition for artificially intelligent applications such as expert systems.

The proposal-in-principle contained within outlines the concept of the research and also describes the likely inputs, outcomes and benefits for your unit. You will appreciate the need to progress reasonably quickly on this matter and to this end I would request a prompt decision in principle. If we proceed with this venture, a detailed research proposal would follow.

I look forward to your early reply on this matter.

Yours sincerely

Simon Ewing-Jarvie

NEEDS ANALYSIS

Simon Ewing-Jarvie met with Sharon Wards on 2 Apr 92 and discussed the challenges facing the Sensory Evaluation Unit (SEU) in regard to the documentation of human expertise. It was noted that steps were already being taken to record knowledge relating to the complex tasks undertaken at SEU. Commercial confidence was raised as an important issue which could possibly eliminate the use of many SEU activities as research vehicles.

Sharon Wards discussed the need for SEU to document the collective expertise of the four resident experts regarding training of panellists. This was acknowledged by Simon Ewing-Jarvie as a suitable and worthwhile research topic which could result in mutual benefit.

OUTLINE OF PROPOSAL

That the SEU agree to work with Simon Ewing-Jarvie during the conduct of his Masterate research. This would involve knowledge engineering techniques, as described in Appendix 1, being employed with three to four resident sensory evaluation experts to model their collective expertise regarding the training of panellists.

OUTCOMES FOR SEU

- * A structured analysis of resident expertise in panellist training.
- * A documented model of that expertise.
- * Publication opportunities for personnel involved.

BENEFITS FOR SEU

- * A greater understanding of the human processes involved in panellist training.
- * Smoother induction of new staff through being able to examine current processes on paper.
- * A reduced likelihood of losing corporate knowledge through staff turnover.
- * Enhanced public image through the use of advanced human resource techniques.

INVESTMENT BY SEU

The investment required to achieve the outcomes and benefits described will involve no monetary component. SEU would be expected to make the expert group available for data collection and validation during the period May to September 1992 which would involve approximately 20 hours per expert.

POSSIBLE TIME AND ACTIVITY SCHEDULE

30 Apr	Agreement in principle.
May-Jun	Initial data collection. Panel training.
Jul-Aug	Data analysis involving expert input into validation.
Sep-Oct	Write thesis and publish joint articles.

Appendix B

Full Transcripts

Full transcript Expert 1 - Session 1

1 KE

2 Thanks for agreeing to participate in this series of interviews Expert 1.
3 What I'd like to begin by doing is getting a general feel for the field of
4 work that you're involved with here. This will enable me to gauge the
5 scope of the subject and establish some boundaries. I will be trying to work
6 out exactly what it is that you and the other panel leaders do when running
7 a sensory panel

8 Expert 1

9 OK

10 KE

11 Perhaps we could start by you giving me a general description, in layman's
12 terms of the panels and their activities?

13 Expert 1

14 Sure. We get them straight off the street. We select them on the basis of
15 how good they are at tasting a particular product and in order to do this
16 we employ threshold testing which is basically how good they are at the
17 four basic tastes. We need to find out how good they are at discriminating
18 between certain products for example a cheese panel gets to try all cheese
19 products. We also use triangle tests where 2 samples are the same and 1
20 is different. They need to be able to spot the difference. At the end of 1
21 to 2 weeks we eliminate about 75% of people. That usually leaves about
22 15 to 20 people. Now some of these people are good at one product but
23 not at another. Then in the first few training sessions after selection we get
24 them to describe everything about the product for instance flavour, texture,
25 colour. We take all the words that are used and the group leader finds
26 matching words to get a shared vocabulary. That usually takes a couple of
27 days. Then we introduce intensities. The group leader works out some end-
28 of-scale words like weak, slight, strong, and then moves from those onto
29 numbers. We would probably spend 2 to 3 weeks on words. In actual fact,
30 the process of converting words to numbers is easy by tying values to
31 specific words. Then the pressure goes on the panel because this is when
32 the leader sorts out the outliers. We can spend 1 to 2 months getting the
33 numbers right. We also start measuring variation, analyzing the standard
34 deviations and giving that feedback to the panellist. At the end of this
35 process we make the decision as to whether this panel is fully trained or
36 not. This is a big decision. Once a panel is trained, they are 100 times
37 more sensitive than a person on the street to that product. We've got to
38 try and get rid of the variation between the experts at that decision point
39 as it's critical to the quality of the panel

40 KE

41 Have the group leaders been given feedback about their variation in the

12 same way that the panellists do?

13 Expert 1

14 No. The impression amongst the experts, I think, is that they're all the
15 same. There's no doubt in my mind, however, that this is not the case as
16 there are definite differences in assertiveness and how hard each one
17 pushes a panel

18 KE

19 I see. What is the general feeling amongst the group leaders toward this
20 knowledge acquisition project?

21 Expert 1

22 Overall, it's positive. I think they're a little unsure about what's expected
23 of them but all see this as part of the overall quality programme

24 KE

25 Good. It sounds to me like the training process is open-ended, time wise?

26 Expert 1

27 Yes. It used to be 3 months with Establishment 1 staff using laymen who
28 are not familiar with the field take longer, as do larger numbers. There are
29 standard deviations for each product based on historical data but leaders
30 aren't using them so my question is what are they basing their decisions
31 on? You see the length of panel training ranges at present from 4 to 6
32 months, depending on how motivated and keen they are. It also depends
33 on how good the leaders standards are and how pushy they get

34 KE

35 Right. So how many hours training per week do they do?

36 Expert 1

37 3 times per week. In training that is a 45 minute session ranging up to 1
38 hour

39 KE

40 And at what point do you start paying panellists?

41 Expert 1

42 During screening they get \$5 per day plus a \$5 bonus to be there at the
43 end. Normal pay for training is \$10 per session regardless of the length of
44 time that they're there. I should point out that it's reimbursement, not
45 payment, that's to get around the tax problem for some. So if there is no
46 panel activity for a while, some find it hard financially

47 KE

48 So it sounds like an expensive business for the sensory evaluation unit

9 Expert 1

10 Yes it is and so we continually have to sort out the non-performers

11 KE

12 Mmm. What causes the decay in performance?

13 Expert 1

14 Lots of things. Changing lifestyle, pregnancy, stress and tiredness. Fatigue
15 and adaptation due to over testing. With chocolate the chocolate data,
16 Expert 2 plotted all the standard deviations. Some major differences
17 occurred but we hope this will change with computers in the booths
18 because they will get instant data analysis and feedback. When we give
19 results back, we hide the names except for those who are consistent
20 outliers. What I want from this is something neutral and standard for group
21 leaders to refer to

22 KE

23 Right. What I'd like to move onto now is to look specifically at what the
24 group leader's tasks are throughout, from the start of selection to the point
25 where they declare the panel trained and ready for project work

26 Expert 1

27 Well what I do at the start of training a panel is different to what I do at
28 the end. It all starts I guess with interviewing. The first interview that
29 you're ever likely to do is going to be on the phone and there's normally
30 a standard sheet that you go through, or if they don't want to do it on the
31 phone, you send it out. So the basic information that you've got on a
32 panellist is what's on the sheet like their sex, what they like or dislike in
33 food, whether they're allergic to stuff, how committed they are. When they
34 come in for screening, toward the end of the week we actually interview
35 everyone and we ask them, from another standard sheet, what their
36 commitment is, are they going to get employed, how happy they feel about
37 coming in. What we're trying to ascertain is if they are going to be a good
38 group person, whether they're going to leave us in the lurch 2 months
39 down the track and whether or not they're actually interested in it. We try
40 to find out without being too subtle whether they're doing it for the money
41 only or whether they're actually genuinely interested. We're still trying to
42 work on that interview thing because we've only done it once across the
43 board and we've already found that some of the panellists crap out
44 inevitably and you could never tell that in the interview so we've got to
45 improve that a little bit. So that's the interview part. What I try to do is
46 have 2 or 3 people in our section sit in on the interviews sort of stay in the
47 room and people talk to the panel leader directly so the others just sit in
48 so it's not just one person making the decision based on that interview.
49 And then we sit down and compare notes. One thing that I am concerned

0 about is that personal impressions come across

1 KE

2 Right. After interviews, what comes next?

3 Expert 1

4 I guess the next stage really reflects what I was talking about, sifting
5 through and sorting out the ones to be included. I guess you could call that
6 the personnel selector. It overlaps with the interview but is just a
7 confirmation. Have we got a good group person in there? But the main
8 selection criteria is still the results from their screening tests. It's based on
9 the most objective information that we can get so in combination with their
0 results from the screening tests and the limits set on them, plus the
1 interview, that's how we would select. We tend to select more people than
2 we need because they usually drop out. If we were to select just on the
3 basis of results, we wouldn't get enough. We often drop the standard to
4 increase our numbers

5 KE

6 I see. So when you have this slightly oversized group and are ready to
7 begin training, what role does the group leader focus on then?

8 Expert 1

9 Well that's a multi-faceted issue and it's where the fun begins. You've got
0 to take a whole group of strangers, people that don't know each other and
1 we've got to convert them into a group of people who know each other
2 and trust each other enough to say well I really don't think that it is such
3 and such and feel comfortable with that. So this is where your skills are
4 really brought to bear as a facilitator

5 KE

6 Good, so should we note down facilitator as a task?

7 Expert 1

8 Oh sure, yes. You've got to sort out the people that don't talk a lot and sit
9 on the people that are talking a lot. The first few sessions are really
0 important ones as they set the scene so we've tried to do some fun things
1 and some interesting things. Especially introducing two groups of people
2 like where we are reinforcing a panel and you have to bring the new lot
3 and the old lot together. It's quite a terrifying experience. The old ones are
4 trying to suss out the new lot and the new ones are thinking they know
5 everything and we don't, so we've done a few little things like blindfolded
6 a new one and the old one feeds them a product and they have to say
7 what they think it is. Then they whip off the scarf and introduce each
8 other. You have to be able to read people and know when to push and
9 when not to. One of the harder roles is knowing when to push them. I
0 guess that leads into the next task which is as a trainer.

01 KE
02 Should I write that down as a task?

03 Expert 1
04 Yep. You use your facilitation skills to get them working as a group and
05 then you use your training skills to know when the right time is to put the
06 pressure on. If you get it wrong you can have a real mess. I've seen it
07 happen where the panel leader let facilitation go on too long and they
08 were having heaps of fun but no work got done. And then there's the other
09 side where the group weren't ready to be pushed and they just fell apart

10 KE
11 I see. How do you tell how they are performing?

12 Expert 1
13 That's a reasonably straight forward thing. If you know that you've got a
14 panellist that's not performing you may not know why. So you've got to
15 analyze the performance, looking at questions like why is the person doing
16 things wrong, what can I do to improve them, what steps do I need to take
17 to ensure that they do pull up, and if they don't making the decision to get
18 rid of them. It's analysis I guess.

19 KE
20 Is that the task you'd like to record?

21 Expert 1
22 Yep. It also comes in when you want to throw a bit of stats around the
23 place. You want to try and link something that's happening with the panel
24 with something chemical or physiological or whatever but analysis comes
25 further down the track. It's when things get tough that you try and get a lot
26 more discriminating and analyze a lot more

27 KE
28 I see. What comes next after analysis?

29 Expert 1
30 Well I guess the panel leader has to make a final decision about the panel
31 or the panellists. We say that the panellists have got this apprentice period.
32 Now I could be really hard nosed and say right, if you're not performing
33 in 3 months, then you're down the road but lots of other issues come in
34 and effect that. Like we have to keep persevering with useless ones to
35 keep the numbers up or they are going to bad mouth us if they leave. It's
36 not pleasant getting rid of people. So there's those decisions to make like
37 who to keep and who to carry on with

38 KE
39 Right. So are you suggesting that final decision maker is a task of the
40 group leader?

01 Expert 1
02 Yes it is. There's also decisions like are they ready now or are they not.
03 Other issues that affect that are if the project work is coming, or getting
04 nagged to make some money out of the panel. It's how the decisions are
05 being made and when the rules are being broken that I want sorted out
06 and the more objective data we can get to help that decision the better

07 KE
08 So having looked at all that, it's clear that the focus for this projects needs
09 to be tighter than just training a panel

10 Expert 1
11 Yes, I agree

12 KE
13 How would it be if we focused on the stage covering when a panel is
14 considered trained?

15 Expert 1
16 Yep, that would be good

17 KE
18 Can a person that is socially disruptive bias the tasting?

19 Expert 1
20 Definitely, both in the results of the panel overall, and also in the
21 motivation of the other panellists

22 KE
23 I see. What I'd like you to consider now are some of the key items that
24 make up the tasks we have discussed. If we can look first at the task of
25 interviewer. What do you feel are the elements that make up this task?

26 Expert 1
27 Well, the most important ones would be... articulation, the ability to speak
28 well to the applicants. Social skills are also important and we also have a
29 checklist that we work to

30 KE
31 Good one. I can see that you understand what I'm after. Can I get you
32 then to think about the items which make up each of your 5 tasks and
33 when we next speak, we can refer back to them then

34 Expert 1
35 No problems

36 KE

17 Just to change track completely. How do you as the section head see the
18 information that is coming from this project being used?

19 Expert 1

20 Well what I want the information to be used for is mainly as a training tool
21 so there can be some consistency and objectivity into the work. At the
22 moment if you go anywhere in the world or in the industry, everyone is
23 doing their own thing. Sure they follow some basic principles but it still
24 relies on the individual and I don't want that. I would rather have it that
25 there are certain rules that apply so that in 3 years time, someone new can
26 come in and train a panel in exactly the same way. So I want some good
27 basic rules to come out of this that everyone will be guided by. So I see it
28 as a key training tool and also as a key feedback mechanism for the quality
29 system

30 KE

31 Mmm. So it sounds like you want a prescriptive model for panel training?

32 Expert 1

33 Yeah. I also see it as an opportunity to change the way we do things. I'm
34 hoping that by being asked to write down what they do, people will
35 challenge the old ways. I see it evolving

36 **DURATION 44 MINUTES**

57 Expert 1 - SESSION 2

58 KE

59 Okay so since now that I've managed to go right through the first round
60 of interviews what's become apparent is that the scope of the term training
61 the panel is huge and you might recall that we talked about tasks, it began
62 like training the analyst and those sorts of things and so in order to make
63 it manageable so that we get a distinct outcome and so the outcome that
64 you most want what I would like to do now is focus on the actual activities
65 that lead up immediately prior to declaring the panel trained and the
66 decisions that go into actually releasing them for work on their projects.

67 Expert 1

68 Right

69 KE

70 So first is we just start with you describing in general terms what it is you
71 do in those last few stages in how you decide.

72 Expert 1

73 Well, I had a time frame of 3 to 4 months at Establishment 1. Because
74 that's the only sort of experience I've got we used to three or four months
75 so after they've been going sort of two or three months you used to start
76 looking quite carefully at their individual results, ok. So we would actually
77 start using a calculator or putting it through the computer often we had at
78 the time and getting out the mean and getting out the standard deviations
79 and having look to see if that standard deviation is coming down. Now
80 that's how we used to do it scientifically but as I have talked to you before
81 some of it goes on gut feel, you know with the way the panel is talking or
82 the way that they are getting the scores or the way it is showing up on the
83 white board that the variation is coming down. So, what's in the
84 development of the QA manual I've been trying to get them to use
85 standard deviation as the guide but in the past it often used to be the gut
86 feel, looking at the scores and thinking they are starting to come in now.
87 But if they are beyond hope then you would have got rid of them but if not
88 you just put up with what the results were. So, it's really on the numbers,
89 it's a numbers game now and sometimes a project work is about to get
90 dumped on you that will help make sure that it is all well. So, if you are
91 given plenty of time it would be based on the variation by low standard
92 deviations, but if the project work is at fault you might push them a little
93 bit more and make that decision perhaps earlier. The way that the sort of
94 process we go to just before we consider them fully trained we do what we
95 call half blind tests, where we take them through the training situation, put
96 them in the booth just to get them used to that then after they have done
97 the testing in the booths you take them into the round table room then
98 and you right their results up on the board. It's sort of half and half the
99 training session half and the actual tasting session so it sort of gives them
100 time to practice that transition from being in the round table room all the

01 time to actually doing the tasting in the booths, and that sort of gives them
02 a bit of confidence as well in their results and doing it on their own they
03 want to know how they have related to the other one's in the room. So we
04 probably do two or three weeks of that on perhaps a bit of a mixture of
05 project work and a mixture of standards to see how the variations have
06 gone. And then if that hasn't worked out well you give them another shot
07 at the training but normally by that stage we've got it pretty well sussed so
08 that's a rough outline of the sort of process but it depends on standard
09 deviation, on your gut feeling and on how the project works. Those are
10 the sort of decisions around. Sounds very airy fairy which is why I want to
11 try and get it a bit more cut and dry. And also consultant, you consult, you
12 talk to the other staff the other research officers and say well you reckon,
13 and sit around and look at the results and the odd days we used to do an
14 analysis of variance on them and everything so give an outline and whether
15 or not they were within the standard distribution and stuff like that so
16 there is a bit of analysis going into it, perhaps not so much down here
17 other than using standard deviation.

18 KE

19 So is standard deviation always used as the criteria for

20 Expert 1

21 It is now because of the QA procedures. Well that is the thing it is meant
22 to be used but I notice some of them, I mean I have a go at them cause
23 you know issue standard down and some of them aren't happy that it is
24 down low enough but because of pressure on them to get the project work
25 done then you have to make compromises, but I have said to them as long
26 as it is reported that the standard deviation was or is as low as we wanted
27 it to be then we don't have a problem but we've got to keep them at
28 thinking standard deviation because that's the only measure that we've got
29 at the moment - the only logical measurement other than gut feeling.

30 KE

31 So this decision, is that based on one or more half hour sessions?

32 Expert 1

33 It is definitely based on more than one. See some of them likes of Expert
34 3 she's been running a lot of half blinds. Normally we probably run two
35 or three weeks it really depends on how good they are, if they're not any
36 good well you whip them back in and work them back in the half blinds
37 again but whilst they're not two or three weeks, it's just until they feel
38 comfortable working in the booths because it's such a culture shock to
39 them they go in there and they have got no one to hold their hand, and no
40 one to cheat off, and they are sitting in the booths all alone its really just
41 to get their confidence up and assure them that yes when they look at their
42 results on the board they are still within, still even with the rest of them.

43 KE

14 So you are looking at six to nine sessions, where you recall the evaluation
15 phase at the end.

16 Expert 1

17 Yes, you're really evaluating how good they are and it's just a chance, we
18 usually have a look at them and then get used to the actual project work,
19 it would be about two or three weeks I suppose KE. That is the way I do
20 it.

21 KE

22 And what is the standard deviation that you are looking for?

23 Expert 1

24 That varies depending on the product, a couple of years ago we did a
25 study, because the literature doesn't give you much guidance on how to
26 check panels variance, it was really a measure that we decided on that we
27 could effectively use that was efficient and it was something that we could
28 do easily so we did a study, we got a summer student to collate all the
29 results from every product that we had and work out the average standard
30 deviation because there was no point in going for something that was too
31 low or wasn't even reasonable and because we use different scales for
32 different products we had to work out standard deviations based on a line
33 scale and a category scale wouldn't work, so she found that on average that
34 the really average attributes like sweetness or something that they do every
35 day, it could even go below one okay, we found an average standard
36 deviation of .5 for instance for sweetness in milk powder which, I mean
37 that is almost perfect given a group of people. But some of them were
38 over one on the ten point scale so you know you're looking at ten percent
39 fuzziness in the data which is quite reasonable really, so it depends on the
40 attributes, if they are not very common attributes then they are very
41 difficult doing and standard deviation might be something like 1.5, I get
42 very nervous if it goes over that because that is twenty percent error in the
43 data which is not very good, I mean I wouldn't buy a machine that had a
44 twenty percent error in it. So the standard deviation will vary on the
45 attributes and it is up to the panel there to decide given the history that
46 he/she knows about the panel is actually what that standard deviation will
47 be, and we are finding here that the standard deviations are a little bit
48 higher than what we would have expected out of Establishment 1. So we
49 are still working through that but at the end of the day those ones who are
50 doing project work have a pretty good idea what the standard deviation
51 should be for most of the attributes. So it is really defining those attributes
52 and standard deviation and getting it whacked into the QA manual. It will
53 probably change depending on the project, if the project work is not very
54 variable then the standard deviation should come down, it's when you get
55 samples that have got such a wide range than any project that you get a lot
56 more variation, so, it's really used as a guide we try and report so people
57 know the sort of quality tool that we are using and if those standard
58 deviations are continually high then they should be shoved back into

39 training, and given the ground table. But, it depends on project work, so
30 we have got our ideals, but in reality it may not always come up to that,
31 which is a problem.

32 KE

33 So, are you looking at the standard deviation of each individual attribute
34 passing the acceptable threshold or the average say of the product as team
35 attribute, you just look at average standard deviation?

36 Expert 1

37 No, we are looking at eventually we should have standard deviations for
38 every attribute, but that will only happen with history, so that we have got,
39 see what she did was she calculated standard deviation for all the
40 sweetness scores ever in milk powder and the average standard deviation
41 was .5 so that is what we have used, so we have used history to give us that
42 information not thinking of an ideal and using that, because we have just,
43 well we are beyond it all of the time.

44 KE

45 Right, so this will be evolutionary in that a panel might generate an
46 attribute called mulgyness and that starts from scratch.

47 Expert 1

48 Yes, exactly. So I mean some of the new panellist that are trained at the
49 moment, they are coming up with words that we have never used before
50 so we have got no history in which to gauge standard deviation but
51 depending on the scale KE, like I said if it get any where near two on a
52 ten point scale it is a problem no matter how much history you have got
53 to base it on, so it's any where near .5 and 2, any of those attributes should
54 otherwise that tool is not fined tuned enough no matter how much history
55 has gone beyond. So, my gut feel is if the attribute is always causing a
56 variation I would get rid of it, because it is obvious not enough people are
57 using it and can relate to it in order to get that variation down, so that
58 would be one of the decisions I would make, any problem areas I would
59 get rid of them. I am a bit tough like that.

0 KE

1 Okay, so the testing and the threshold is panel dependent?

2 Expert 1

3 Yes

4 KE

5 Is that product dependent or something else, some other attribute of the
6 panel?

7 Expert 1

28 I would dearly love it to be product dependent, the problem that we have
29 is that we don't actually have enough history panels as well, I mean if we
30 start a milk powder panel up here that will be the third milk powder panel
31 that we have had in history, now I would dearly love those standard
32 deviations to be the same as what we have calculated with those other
33 ones. Now whether or not it is going to be because we are using members
34 of the general public I don't actually know, I would dearly love to aim for
35 that and then at least we have got some consistency. So at the moment
36 KE my gut feel would be panel dependent but I would aim to have it
37 product dependent at the end of the day.

38 KE

39 What is it that's varying if not the product?

40 Expert 1

41 Well this is really why you're here. Part of me thinks it's the panel leader,
42 part of me thinks it's the mixture of people that we've got, OK we've done
43 our best to try and make sure that they're all similar in their discrimination
44 but, I mean, if you sat in on the six panels that we run, all of them are
45 different. The old chocolate panel won't say boo to a goose and I mean
46 that is going to affect how they , you know, change their scores or be
47 influenced by other people. Some of them, I mean baked products panel
48 you can't shut them up. So, I really think that it's a mixture of panel
49 leader, panel and also probably the variation in the product, and if you ask
50 me say what's the contribution of each of them, I couldn't tell you. I mean,
51 that's something really fundamental and really interesting that we should
52 be working on but given the sort of workload that we've got it's not
53 possible but it's a fascinating area which I'm sure that no-one else has had
54 a look at either because you just wouldn't.. close your eyes and..

5 KE

6 ...it's too big..

7 Expert 1

8 Yeah, that's right, I mean we're only touching a tiny little bit of it...and you
9 know there's lots..I mean psychologists would have a field day out here
0 trying to work it all out. Yeah, that's why we want to try and get at least
1 one part of it cut down to size that we can actually work on so that's why
2 we chose the standard deviations, we could have chosen any other tool I
3 suppose but at the end of the day, at least it's easier to measure than
4 something else.

5 KE

6 What system do you use in the panels then for the panellists to record and
7 report the scores about the product in use?

8 Expert 1

9 Well you've seen the questionnaires that we use haven't you? You've seen

70 them.

71 KE

72 No

73 Expert 1

74 Well I mean a panel sheet for all the different products.

75 KE

76 Oh, I've seen some of those

77 Expert 1

78 Yeah, yeah well that's where they score their things so what are you asking,
79 you asking...

80 KE

81 There seemed to be a lot of different ones and I was just shown a selection
82 so which one is it that you use most of the time? ... or is there no set one?

83 Expert 1

84 No there's no set...well what happens in the training process they start out
85 with a blank sheet, they move on to a sheet that's got a lot of words on it
86 and those words are eventually cut down to what we call our final profile.
87 Then it's on those words that they do their scoring, be it on a line scale,
88 which is chocolate at the moment, be it on a nine point scale which is
89 AMF, be it on a ten point scale which is most of our products, category
90 scale. So your panellists are either using numbers or they are using a line
91 scale with which to do their scoring and it's from that that we then convert
92 that into a mean score with a standard deviation so at the end of the day
93 the panellists don't get instant feedback as to what the standard deviation
94 is on that day but that will change once we've got our computers in, that's
95 the whole idea, instant feedback. If you can say to a panel that day that
96 they're way out of kilter, I mean two weeks later it's no good hearing that
97 news they can't do anything about it then, which is what's been happening.
98 It takes us that long to get it analyzed and get that feedback, and each of
99 them have different ways, Expert 2's probably a good one to ask that
100 question they have different ways of showing how variable their panel is,
101 giving them their feedback you see, and Expert 2 draws little pictures and
102 Expert 3 will draw little pictures, just to show them how out of kilter they
103 are but it's not instant feedback KE so you can't change what's gone on
104 that day which is the sad thing so but the one particular panel, Expert 2
105 for instance did a whole investigation of panellists variability across the
106 year and you could see the standard deviations going up and down
107 depending on the project. The project that had very little variation, the
108 standard deviations came down but as soon as the samples get really
109 variable it throws them out completely and that's when the standard
110 deviation goes up, so with all the little, different ways of giving the
111 information on that variation back to the panel. And then, I mean, I don't

12 do it. It's something that we talk about but they've all got their different
13 ways of doing it.

14 KE
15 You don't give standard deviation feedback at all?

16 Expert 1
17 No..well I don't have a panel.

18 KE
19 Oh, I see. Why do the group leaders, why do some use a line scale, some
20 use a nine point and some use a ten point scale?

21 Expert 1
22 I knew you were going to ask me that. There's lots of theories KE, as to
23 scale use OK, and the psychologists will say that...see some people have an
24 aversion to numbers. Any scale that you use, there will be end-of-scale
25 avoidance, so your idea is to try and get a scale that's wide enough to
26 cover all possible options but also account for end of scale avoidance. So
27 If you've got a five point scale you're only effectively going to get three
28 numbers used because everyone will avoid one and avoid five. So that's on
29 of the aims of our ten point scale, that has enough, gives you enough
30 discrimination and, for them to avoid zero and ten but you're still getting
31 enough discrimination on the rest of your points that are left. Now Expert
32 3 uses a nine point scale, I'm not quite sure... her and I have had long
33 discussions as to her using a nine point scale, so you should ask her about
34 the theory behind doing that. A line scale is meant to get rid of this
35 apparently because with the category scale like a number scale the
36 assumption is made that the difference between one and two, and the
37 difference between two and three is the same. It's not actually true. People
38 can't, when you're talking about intensities of flavours, it's not necessarily
39 linear, and that's one of the problems with the category scale. With the line
40 scale apparently the perception's different. It's that people don't think
41 numbers, they think, I don't know how you'd explain it, see there's lots of
42 literature explaining, doing the comparisons between the different scales
43 and for different projects we decide... with chocolate they all want to be
44 very finely discriminating tools see. And the line scale is used because it
45 had been used by Establishment 3 in Australia, it had been used by Lord
46 Zukerman in England. So we used it to keep that consistency going but at
47 the end of the day it depends on the project, what sort of scale we would
48 decide on. Expert 2's the only one who is using the line scale at the
49 moment here.

0 KE
1 So you can get percentage results on a line scale?

2 Expert 1
3 Yeah, what we've used is what we call a 150cm line, and we have intervals

54 at 10mm and 140 or whatever it is. And then, so you work out in terms of
55 150cm but, yeah, you convert that basically into a percentage.

56 KE
57 And I suppose they can put it on any position.

58 Expert 1
59 Yeah, any position

60 KE
61 Whereas, I guess on a nine or ten point scale they are taking the actual
62 numerical, or integer points only

63 Expert 1
64 That's right, in that case. So the line scale is meant to be more
65 discriminating but I mean, that's my point of view, you get a different point
66 of view from someone else depending on which one they're in favour of.
67 But history had it that Establishment 1 they just used to use a ten point
68 scale and that was it so we're trying to change a lot of history as well by
69 choosing different scales.

70 KE
71 If you are looking at historical data being your threshold what in-built
72 variation is there if you introduce data from a nine point scale and then
73 some from a ten point scale?

74 Expert 1
75 Exactly. History has it that the only thing we've ever used with chocolate
76 is a line scale so that's OK, but I know exactly what you're getting at, that's
77 been one of my arguments with some of the new scales that Expert 3, for
78 instance, has introduced. Luckily, she's only ever introduced it on a new
79 product that we've never done before but that's my argument that when
80 you start interchanging the scales with the product we're going to run into
81 problems with the standard deviation. And what we'll have to do at the
82 end of the day is OK, on a ten point scale it might have been 1.5, on a
83 nine point scale it's got to be lower otherwise you've got a lot of explaining
84 to do. I think there will be problems but also it may not be that big a
85 problem because no-one looks back at old historical sensory data. None of
86 the scientists do and so the relationship between something that was five
87 years ago and something that happens now, you know, it's pretty airy-fairy.
88 We've only been using it for a guidance tool for us, to give us that
89 information. If we find over storage time, if you change the scale, you're
90 up the creek without a paddle and we've had that problem once or twice
91 and I've had to jump on them. You have people come in wanting to
92 change things well you can't do that in the middle of a storage trial
93 otherwise you get no relativity. It's like changing the panel after our first
94 storage trial, the problems that we've had have been just a nightmare.

05 KE

06 So you'd be happy if the same panels are using the same scales over time,
07 that would be acceptable.

08 Expert 1

09 Yeah, it would be sufficient, as long as the panel using the same scale over
10 time, I would be happy. OK, I can see variation using different scales with
11 different panels but there has to be some consistency within that panel.
12 Otherwise, if you use the same group of people you're trying to change
13 them, trying to change them from thinking ten points to nine points or to
14 line scale or whatever. It's actually going to become an issue because a lot
15 of the demands that we've got mean that they want... the scientists want
16 more flexibility so we are actually going to have to train panels to use
17 different techniques like Expert 4's using magnitude estimation with
18 bitterness but there'll come a time when he's going to do category scale
19 with them and that's going to be an interesting transition to see what effect
20 there is on the result and it's probably fundamental that we look at that
21 transition period and then look at changing them back to standard and see
22 what the differences will be. It's on-going research for us to look at these
things but .. there's a lot of literature comparing different methods but at
the end of the day it's gut feel sometimes which makes you make these
decisions. But the thing I find amazing KE is that even though I'm not
doing it, I can still walk into any of those panels and pick up where that
person has left off. It's that in-bred in all of us that anyone can pick up
where the other person left off. Yeah, it's just amazing those gut feels in
the guidance rules that you come up with, are in-bred into all of us so but
I don't want it in-bred, I want it to be written down. So that if we did all
get run over by a bus, you could come in and take that panel. You could
be our next staff member KE, on contract...

23 KE

24 Most of the data collection systems seem to be quantitative, but we know
25 that there's a body of people in the world who aren't quantitative, they're
26 not left-brain operators at all. I'm curious as to what actually happens with
27 them because in a sample of people walking in off the street, a percentage
28 of them will be...

29 Expert 1

30 Exactly, I know, they're our problem. On any sheet that you look at, there's
31 room for comment and these qualitative people will always write
32 something...

33 KE

34 The right brain people...

35 Expert 1

36 The right brain people, sorry, yeah. So that's the option for them because
37 we didn't used to give them that option and they always wanted to tell us

38 something else and at the end of the day, depending on the project, we will
39 take into account those comments because sometimes they think words
40 rather than score and they'll say look it tasted really something or other
41 and you think why didn't you put the score up there. So we'll often take
42 into account, now we don't actually manipulate the data, because you get
43 fired from Establishment 1 for doing that but sometimes if, when we're
44 interpreting the data, if enough people have made this qualitative comment
45 then we'll actually use this information more than the actual numbers.
46 Which is really quite important because sometimes in a particular project
47 none of the attributes will actually say what their overall perception is, and
48 so you can actually miss out on quite a bit of information if you don't take
49 that into account that qualitative stuff. But because, at the end of the day,
50 we're trying to get an instrument that gives us quantitative information
51 that's why we concentrate so much on numbers. But, that's why it's
52 important that the sensory person interprets the data because it's not just
53 numbers and in some of our reports you will notice that we'll say that the
54 panellists commented that the sample was ... and use just words. And that
55 can often give more information to the scientists than the actual numbers.
56 But it is a real problem and those people can be so hard to beat into
57 shape in terms of the numbers. Some of them will write essays, they'll turn
58 over the piece of paper and... we had one panellist ring up not so long ago,
59 she'd forgotten to write down a whole lot of things so the next day she
60 rang us up to make sure that we'd write down on her sheet. They're really
61 keen on telling you all about it but if I was doing it I'd actually interpret,
62 I'd use that information as well, especially in a project that we didn't know
63 much about. So, I know we don't leave a lot of room for it but the
64 opportunity is there now because people want to say things even if you
65 ignore it, they still want to say it and, I mean, that's important that they get
66 their chance to.

7 KE

8 I guess if you were being really hard nosed you'd do the whole thing more
9 efficiently without them, just use really quantitatively minded people who
0 can turn sweet into ten just like that in their mind. So what steps are there
1 if any in the selection process to actually filter out...

2 Expert 1

3 You can't do that with the selection process the way we've got it designed
4 at the moment because we're not looking for numbers ability then. It's
5 only, and that's why we've got them on that apprentice period, it's that
6 some of them just cannot get the numbers game sorted out. We had one
7 chronic panellist out at Establishment 1, on a ten point scale and the
8 highest she would ever put was three, no matter what you did to her. She
9 just had this mindset that everything was very low and even if you gave her
0 the grossest sample that everyone else was putting ten for, she'd still only
1 put three. What do you do with someone like that? OK, she'd contribute
2 qualitatively, she'd always write something but it was just such nonsense
3 and it just put so much variation into your so-called tool that she was just

absolutely useless but because she was out at Establishment 1 we couldn't actually get rid of her. You had to accept the fact that she was a panellist and live with it so, in the end we stopped asking her. But yeah, there are people and the selection process we've got at the moment doesn't allow us to screen them out. It's only in the training process that they're avoiding ends of scale, not using...you know they might only use 3,4 and 5 ever, that's when we start to cut and parry.

KE

Have you ever used psychological testing with a panel to try and identify these personal differences?

Expert 1

I'd love to. I don't know how we'd get on with the number of people we'd be looking at. But yes, I mean you've probably seen that we start out with an interview so we are gradually increasing the.. it might turn into a four day event for these panellists by the end of, you know, the next two or three years but it's possibly worth looking at.

KE

One of the things that we found in doing personality typing that you're feeding information back to the group is an awareness of what's unique about you makes you more, not only more accepting of everyone else but also it enables you to compensate for your differences a lot more.

Expert 1

Right. So would you use Myers-Briggs or something like the four one. What's the name of...

KE

Mok, Paul Mok's Communication Style Survey. Same sort of thing. There's about ten or so tests based on Carl Jung's work. Something like that is probably the most widely used and relatively quick to administer and to feedback.

Expert 1

Right. It's probably worth looking at KE, it's just finding...cos, yeah, we have to stack a lot in that week in order to keep... We cancel all the panels for that week so other project work is affected whenever we do any training you see. So whether or not we administer that as part of or after we've semi-selected them or something, do it to an even wider group and cut them down after.

KE

Yes but any personality test is invalid as a selection mechanism because it doesn't predict future behaviour so I wouldn't see it as being valid as part of the selection system but as part of the training system that just makes them more aware of their own self. In the same way that feedback

5 standard deviation.

6 Expert 1

7 Yes, it would be quite interesting actually. Be a real mixture.

8 KE

9 Oh, for sure.

10 Expert 1

11 But would it be better given the sort of things that we're after to have a
12 mixture or would it be better to go for one or two particular types.

13 KE

14 Oh, I would think that you're never going to get away from the ideal. The
15 ideal is a straight slice of society. The whole idea of personality testing is
16 simply to highlight preferences not to say right we'll a certain type here.
17 Otherwise you sort of end up with a nazi ideal panellist...they all have
18 brown hair and blue eyes.

19 Expert 1

20 Yeah, we can do without that

21 KE

22 You would have quite valid attacks being made on your selecting.

23 Expert 1

24 Yeah that's right

25 KE

26 Anyway, so that's covered that. Just going right back to the start then when
27 we talked about coming into that 2 or 3 week period where you were doing
28 the half-blind tests. Are there any other techniques that you would use to
29 test a panel's readiness?

30 Expert 1

31 Not that I can think of. You've got a pretty good idea from the building up
32 the fact that the numbers are starting to come into line on the white board.
33 Your gut feel is telling you that they're working well, they're still well
34 motivated, they're ready to get on with something. They always keep
35 asking, when are we fully trained? that's probably about it from my point
36 of view KE. But yeah, we've got to get...at the moment the only tool we've
37 got is standard deviation. At the end of the day we've with copy sets we
38 might find that a certain F ratio or significant score is what we use as well.
39 But it's very difficult at the moment. I imagine that it will get a bit more
40 sophisticated but ... till we settle down, it's what it'll be like I'd say. That's
41 all I can think of.

42 KE

So the indicator for you to start considering the final testing phase is a feeling, watching the figures.

Expert 1

Yep, a feeling, at the moment it's a feeling and I would strongly advocate calculating standard deviation to back that up. Cos in the past it has only ever been feeling and once the project actually got started and you started doing the stats on it, you know, your mind would boggle if you saw a huge standard... see in the old days we didn't actually used to worry about standard deviation, it just used to do it and I don't think that's good enough for scientists... scientific research so that's why I'm on my orange box about the whole thing cos we've never had these sorts of tools in the past.

KE

So what... you said 2 to 3 weeks, what is the minimum number of sessions with an acceptable standard deviation that you would say OK, that indicates to me that...

Expert 1

If they went a whole week with an acceptable level of standard deviation and I'd given them a wide enough range of samples to cover most of the attributes I'd be happy with them then and if 2 weeks prior to that they were all over the show then I might even extend the half-blind period.

KE

Right, so 3 sessions...

Expert 1

3 full sessions of really good standard deviation on a wide range of products, I'd be happy.

KE

And, how many samples for each attribute would you say is a wide enough range?

Expert 1

Well, you need to...I'd probably throw in a control every time, something that's just standard so that you're actually testing their consistency cos you need to find out... if you threw any duplicates in then they should actually score it the same. If they didn't score it the same, then you'd be concerned so you'd always throw in one reference every time and I'd throw in, for the standard attributes like sweetness, creaminess or whatever, I always have something that is representing that every time. And then I'd perhaps throw in one or two odd ones like a really high sour one or a really high oxidised one or something like that. It will depend on what you've got on hand but that's the procedure that I'd take. So I'm really looking for...I'd say 50% or more are standard type samples you know routine stuff cos that's mainly

14 what they would be doing and the rest of it just one off type stuff of any
15 other type of stuff they're likely to encounter.

16 KE

17 Is that at each session or over the whole...

18 Expert 1

19 Over the week, yeah. See the problem that you have KE is that if you
20 throw in... and that's something else you have to take into account...see if
21 you have three really bland samples, you know really high oxidised ones,
22 you are going to get a higher score than if you had put 3 or 4 oxidised ones
23 together, so you have to take into account contrast bias as well. That's
24 what the official word for it is, so that you've got to be careful that you
25 don't bias the standard deviation, I mean, we don't go in there and say,
26 right I want a really good standard deviation so I'm going to make this one
27 stand out so that everyone will get it. So you've got to do a bit of a mix but
28 normally I would try for something, say a reference and then sort of build
29 it up in intensity so you've got a little bit of oxidised, medium oxidised and
30 high oxidised or something, and they're all randomised anyway so they're
31 not all going to get them in the same order. Something like that anyway
32 cos you've got to think of that contrast bias. That's a bad one to fall into.

3 KE

4 Is there a point where you think, I've made a mistake, I shouldn't be doing
5 half-blinds at this stage because their performance isn't there, I'm going to
6 scrap this and go back to the training? How many sessions before you'd...

7 Expert 1

8 Oh god, if I'd gone 2 weeks and couldn't see any improvement I'd say
9 scrap this, let's go back.

0 KE

1 So 6 sessions?

2 Expert 1

3 Yep. I'd say anywhere between 6 and 10 sessions that's when I'd make the
4 decision. Cos, yeah if they're not capable at that stage then they weren't
5 ready to go in in the first place. So your gut feel or your standard deviation
6 calculation was wrong.

7 KE

8 Right, and you go back to the round table?

9 Expert 1

0 Go back to round table training and you know, beat them round the head
1 a bit more.

2 KE

3 Not literally?

4 Expert 1

5 Your poor transcriber is going to think there's some strange things going
6 on down here.

7 KE

8 And what sort of activities will you take them back to as remedial
9 activities?

0 Expert 1

1 I'd take them back to standards, back to all the basics, what you'd used as
2 training for sweetness, what you'd used for each for each of the attributes
3 to get that scale delineated again because at the end of the day, we should
4 have references, say on a 10 point scale, we should have a sweetness score
5 of 0,2,4,6,8,10. And I'd go back and sort out which one was a problem
6 area, which ones they're having the most problems with, so I'd probably
7 spend say 2 weeks on references again, and then try throw some more
8 project work at them again and see if they'd sorted their lives out then.
9 And then shove them back in, and if they keep doing that then you think,
0 I've got a real problem with the whole panel and what's the story.

1 KE

2 Or a real problem with the group leader?

3 Expert 1

4 Yeah, that's right, that's true actually. We have changed group leaders
5 because of that too. It's a high possibility.

5 KE

7 Perhaps, this is one of the potential uses of the psychological tests. You
3 might find a communication style mismatch between the dominant part of
9 the group and the group leader, where it doesn't matter what you say it's
0 not getting through.

1 Expert 1

2 That's a good point actually KE. Cos, I mean you probably would have
3 found that all of us are quite different in our personality and that
4 influences like, I would train a panel differently to Expert 3 or Expert 4 or
5 Expert 2 would because I'd probably be pushing people a lot more than
5 they do.

7 KE

3 The most noticeable difference is the varying levels of confidence because
9 that transmits into assertiveness and expectation and students perform to
0 expectation.

1 Expert 1

2 Yeah, that's right, yeah. The panels that I... I wasn't very confident when
3 I first started 8 years ago but I guess it's just practice and learning your
4 people skills a lot more, knowing when you've pushed them too far and
5 back off, back off otherwise we lose their confidence forever. And it's just
6 a matter of somebody picking those things up, but yeah, I mean, as I said,
7 part of that variation is due to the panel leader and I want to try and get
8 that sorted out and if psychological testing or communication style testing
9 is going to give us a better handle on it then...

0 KE

1 Well I think it will give you another handle on it, it's just part of the
2 picture.

3 Expert 1

4 Yeah, that's right.

5 KE

6 I guess the question that follows from that is, is there a threshold where
7 the group leader knows that they need to withdraw?

8 Expert 1

9 Oh, that's a tough one, that one sometimes they can't actually withdraw
0 and that's been the problem in the section because we have so few staff to
1 shove around sometimes people have been forced into a situation where
2 they can't get out. Now Expert 3, for instance, she used to take the milk
3 powder panel at Establishment 1 but because of problems that we were
4 having with her stress and her workload we actually had to pull her out of
5 that and put Technician 2 on it and, you know, you could see changes
6 almost immediately. And that concerned me, but more often than not we
7 can't change them around, it's just that's whose got it and good luck, see
8 what you can do in the meantime. But, for instance with Expert 2 and her
9 chocolate, she's been really quite concerned. Her confidence in taking
0 panels has gone down so I actually sit in on sessions now and try and help
1 and this buddying system that I've been trying to get going with the other
2 ones as well to ensure that if someone is sick, that the other person has
3 much more confidence going in there knowing what has gone on before.
4 I mean, I don't have a problem going in on any of those panels but the
5 other ones would, you see.

6 KE

7 Yeah, I think it's good sense to have the backup in case someone's not
8 around. I guess my question's more directly targeted toward the situation
9 where the group leader can't achieve the level of confidence or threshold.
0 Because with this huge investment in time in training a panel...

1 Expert 1

2 ... \$30,000...

3 KE

4 Can you afford to declare the panel incompetent?

5 Expert 1

6 No...yeah...

7 KE

8 It's a \$30,000 decision isn't it

9 Expert 1

0 It is a \$30,000 decision which I have to make now. In the past I didn't have
1 to. It's going to happen, I know it is and I'm going to really balk when I
2 have to make it. We have talked in our research meetings about problems
3 that we've had with the panels and that... if they were getting other
4 people's ideas at that time that's helping but at the end of the day you're
5 still on your own in there and I've been called in quite a few times with
6 Expert 3 in particular to try and sort out a panel and try and get them
7 back on track. As you said, whether it's their confidence or whether it's
8 their assertiveness I don't know but I want to try and get that sorted out
9 because that does concern me and if going through this process will
0 identify all of the steps required then all of us can work out together where
1 the areas we're not very strong in and try and get those sorted out now...
2 A real problem... yeah... I mean if we can clone the perfect panel trainer,
3 you know, a hologram or something.

4 KE

5 But you can't of course because the other members of the panel aren't.
6 People click or they don't.

7 Expert 1

8 Exactly. Yeah I notice that with some of those panels that you just haven't
9 got the rapport with them that you have with other ones. So, I mean I'm
0 sort of up to the stage now where I try to change my style depending on...
1 and that's where they say that I'm quite good that I can identify... but that's
2 only through practice you just pick up the vibes that you're getting and
3 quickly change depending on it but yeah, some of them have still got a way
4 to go on that. That's just going to come with training more panels.
5 Something like this should make it a bit faster.

6 KE

7 Certainly, as I've already said, bringing the awareness levels up is going to
8 serve your purposes.

9 Expert 1

0 Oh yeah, God I've never had to think about it so much before you came.
1 If nothing else that's what it has done.

2 KE

3 OK well that's all I had for today. I'll let you off the hook.

4 Expert 1

5 Far out. You're a tough one.

6 **DURATION 48 MINUTES**

Full transcript Expert 2 - SESSION 1

KE

OK, one of the important things that we do throughout this project is to ensure that you all get asked the same questions. Now, coming out of what you say more questions will develop in different ways but what you are going to get asked in some way relates to what I asked Expert 1 the other day as well, and it's not a sort of matter of checking up on people to see if they all come up with the right answers. We'll come up with probably different models which achieve the same outcome anyway. So, perhaps if you could just start by giving me a general layman's description of how you train a panel from the time you first see... first place an ad or whatever that would be a good start.

Expert 2

Right. Taking it from the time we first place an ad well obviously we get people who respond to that ad and we get them in for what's called screening sessions which normally occur over a week. They come to 5 sessions where they will be asked to do a variety of tasks normally includes identification of basic tastes and may include doing what we call threshold testing where they have to taste a series of solutions until they can taste and identify the flavour, but that isn't always done. Secondly it will include normally triangle tests which is where they are given three samples, two are the same, one's different and they've got to identify the odd sample. In addition we normally get them to do some descriptive work where they have little vials which have got different kinds of food in them and they've got to describe the smell or something like that and we also do a couple of other things which are mostly for fun and also to fill in time so they won't get too fatigued. One is listening to a sound tape of different sounds and trying to identify the sounds and we also have one where there's different shapes which have been shaded in to a different proportion and they've got to score them on a scale so we just look to see if anyone has real trouble using a scale but that's more of an interest thing for them than the actual results that we get. Most of the testing is based on triangle tests or it has been traditionally. In the last lot of screening, Nicole and Expert 4 have done a few different things because they're doing something different to what we traditionally do with a panel and so therefore they're trying out some different selection techniques. We also try to look at the person's personality, try and weed out any obvious undesirables and so in the most recent lot of screening we actually had a small interview where each person... just trying to find out what their motivation was, was it strictly money, was it interest, were they a student looking for extra money. What their availability was, if they were an unemployed person, just find out a little bit more about them because in the first year we've had quite a drop off with people finding jobs or shifting or getting students and in this calendar year and all of a sudden their timetable doesn't fit and this sort of thing so, that's basically the screening. Then at some stage we set a level above which we will accept people and below which we won't. That

16 does vary according to the kind of screening that's being done, what the
17 product is etc. Depends how many people we've got to choose from. And
18 from that level we'll choose the people we want. We normally allow for
19 more people than we actually want because we get, we always get a
20 percentage who are unreliable, not interested... a whole bunch of things...
21 and then from that the training starts. In my mind there are two different
22 things that we have to do because sometimes we screening for a total new
23 product that we haven't done before and in other cases, which is what I'm
24 doing at the moment, we have an existing panel which we need to add
25 people to so I have to try and train those people to fit in with the existing
26 one and that's a little bit different to training for something completely
27 fresh and things. If it's a completely fresh product which we haven't done
28 before, don't have an existing panel for, we normally would spend a few
29 sessions, maybe 3 or 4 just getting into word generation so you give them
30 different examples of the product and get them to just try and describe it
31 and try and steer them away from subjective words like bland or nice or
32 yuck or whatever to actually trying to say well you think it's nice but what
33 about it that makes it nice or what about it makes this thing disgusting.
34 Things like that. At the same time in the introductory sessions we go over
35 things like how to taste, how to spit out nicely in public, yeah, and a little
36 bit about what the section is about and what the institute is about. Need
37 to try and give them a little bit of knowledge. Yeah so after those first 3
38 or 4 sessions we'd hopefully have a big list of words that people are coming
39 up with and then we start trying to group them. So, some people may be
40 saying it was 'cocoa'ey and some others might be saying it was milo and
41 someone else thought it was 'chocolate'ey, you know that sort of thing.
42 Trying to group the word together into broad categories and then
43 ultimately agreeing on the one or two words to describe that type of
44 flavour. Following on from that we start trying to develop reference
45 samples for each of those areas so like for sweetness you have to find
46 something that's not sweet right through to the limit of what we could
47 expect that sort of thing, and hopefully some points in between, that tends
48 to make it easier for people. At some point along there we'd start
49 introducing some kind of scale depending on what we wanted to do with
50 them. At present we're using what's called a category scale, don't know if
51 you're familiar with that term, like scoring something from 0 to 10 is sort
52 of like category scale so we might start them off by trying to use words like
53 slight, weak, moderate that sort of thing and then transfer them onto
54 numbers when they've had a bit of practice at trying to put it into
55 categories. Another kind of scale which we use for some panels is a line
56 scale, 150mm long with an anchor point at each end. So with that we'd
57 start them off with that rather than using words. But yeah, once you've got
58 your lists of words, it's basically a matter of developing reference samples
59 for each of those words, making sure that people do agree on what those
60 words mean so thereby developing definitions which we will incorporate
61 into the reference samples. One example say, for chocolate, is sweetness.
62 We did that pseudo-chocolate 'Meadow Sweet' OK, and that is a weak
63 reference fro sweetness and Milky Bar is a strong reference OK, so we're

starting to try and find things which are reproducible and are reasonably consistent for the panel. Then, basically once the panellists can easily recognise the different flavours then we start to teach them how to use the scales. Yeah, that's probably the most time consuming bit. So that's it in a nutshell. Along the way, you should be trying to give them feedback on their individual performance. Most of it is done as group work. Occasionally we get them to do things in the booths, A for a bit of variety but also so that we can see, because in a group of 20 people it's often quite easy to miss, that someone is not really keeping up with the play or they're not understanding or that... especially if they're the sort of person that don't speak out. Quite often, we put them in there just so we can check and have a look at what's going on with individuals. Yeah, so that's it in a nutshell. If, like with the chocolate panel, it's something that we're going to add onto, we do basically the same things but we have to try and guide them a little bit so that we don't end up with 2 different lists of words and that sort of thing and because reference standards have been developed previously we'll be using those and seeing if the panel can relate to them. If they can't, we've got problems. * phone interruption * So we'd also be getting the existing panellists to join that group very early. Basically as soon as they've got over their first couple of weeks, they've just got their words and are trying to put them into groups so that we can make sure that everyone understands the standards and if necessary we will use some of them.

7 KE

8 Mmm. It must be a lot harder to get reinforcements onto a panel than to
9 build a whole new panel... to get consistency.

0 Expert 2

1 A lot harder, because you tend to end up with factions.

2 KE

3 Yeah, the dynamics are disrupted. At what point do you declare a panel
4 void and start again for a product?

5 Expert 2

6 That almost happened for the chocolate panel because, for two reasons.
7 One, they got very low in numbers fairly sudden and secondly that we
8 started to get a really big variation in their data. At the moment I'm trying
9 to sort through that and decide which panellists I'll keep. There is 9 of
0 them at the moment and there is at least 2 that will probably need to be
1 removed. So yeah, I'm trying to make rational decision about that and
2 really identify... and I didn't train the panel originally so it has been quite
3 hard taking them over because it seems to me that perhaps they didn't all
4 agree on some of the definitions or their understanding of them was not
5 clear. It wasn't until we actually got into project work that things started
6 falling apart, in quite a serious way. We've never totally given up on one,
7 that I can remember. Normally only because of the numbers getting too

low. If the data isn't good then that's what we report, that the data isn't good. But normally it's a matter of going back and trying to fix that before going on and doing somebody's project. So it's a matter of training but it isn't really that simple.

KE

How many is not enough for a panel?

Expert 2

Less than 8

KE

8. What's the basis for that?

Expert 2

It's basically just for the statistics. Other centres use different numbers but the institute has always maintained the tradition that 8 is, you know, a minimum to make it sort of valid. I guess the less you have the easier it is for your standard deviation to get out of control. I don't actually know what the actual reason is.

KE

So the number 8 isn't actually based on any particular research, statistically or sensory...

Expert 2

Not... 8 is quite a common number in the sensory field used in research. As far as I know, we've never done a panel, well I never have I should say when I did the same items with 6,8 or 10 people to see what happened. It's a tradition thing, historic.

KE

Way back

Expert 2

Mmm, way back

KE

How big do panels get?

Expert 2

Normally, after screening, we start off with about 20 and they sort of whittle down within the first week or so to about 16 or so. We tend to lose a few. Around 20 to 15 is a good number. We can then get them... if they're going away on holiday we don't drop too low. If you only have about 10 and one gets sick, you can really have a problem. Yeah, around 15 is a good number. 20 is hard to control

5 KE
6 Yeah, it must get pretty crowded around the table

7 Expert 2
8 Yeah, it's pretty crowded and when there's a whole lot of people they are
9 more vocal than ... but I mean again that's something that's reasonably
0 subjective. I know that Establishment 3 in Auckland has got a maximum
1 of 10 or 12 around the table.

2 KE
3 That's their tradition?

4 Expert 2
5 Mmm. They believe that it works better and...

6 KE
7 So, this training process... at what point do you declare the panel trained?

8 Expert 2
9 In theory, we declare it trained when they can reproduce or at least
0 evaluate the same samples, when all the standard deviations are within the
1 limits that we've set and sort of remain there quite happily. A third factor
2 that comes in quite a lot is the pressure to have them ready. So if you've
3 got heaps of people moving up the timeframe, that process does get
4 shortened a bit. And also there's the logistics, the historic logistics anyway
5 of monitoring the standard deviation, I think that hasn't always been done
6 as well as what we should be doing.

7 KE
8 Mmm. Under commercial pressure, who makes the decision, the group
9 leader at the time or the section head.

0 Expert 2
1 The group leader. I mean I would discuss it with Expert 1 but I'm sure
2 what she would say is, if you don't think they're ready, don't do it. We
3 might decide that we'll do a round table discussion but we won't do any
4 statistics or anything if there's real pressure on. That normally is the group
5 leader's decision

6 KE
7 And you mentioned the preset standard deviations, where do you get those
8 from?

9 Expert 2
0 It's only something we've done recently. We had employed a summer
1 student last summer and she, for existing panels, went back through whole
2 heaps of data and worked out what the average standard deviation was for
3 the different attributes and then we tried to set limits that we thought were

4 reasonable and achievable but not so easy to reach that they were
5 nonsense. Most products or most attributes it's plus or minus one on the
6 scale but some of them go up to 1.5, for attributes that we know the panel
7 have more trouble with. So yeah, you've only got a 10 point scale so when
8 the standard deviation starts to get too high it sort of makes a nonsense of
9 it, of the data.

0 KE

1 And the... so you are building an historical database which basically
2 controls future activity?

3 Expert 2

4 Yeah, I mean it depends on the type of scale and that sort of thing.
5 Basically you want to show that the panel is able to score using the same
6 part of the scale and can do it consistently.

7 KE

8 So would I be able to get a copy of that data or those standard deviations
9 for the products? It will be interesting later on for instance when we start
0 to look at the actual numbers involved.

1 Expert 2

2 Mmm

3 KE

4 How long does it basically take looking at the longest time and the shortest
5 time to train a panel?

6 Expert 2

7 Well a really simple product I would say 8 weeks is really the shortest,
8 that's with them coming to 3 sessions a week. And the longest... oh 6
9 months maybe.

0 KE

1 What sort of product leads to that complexity?

2 Expert 2

3 I think that things that we've trained one for that long here have been
4 things like baked products. I anticipate that chocolate is going to take quite
5 a few months. 3 is sort of about the average. Things like butter and whole
6 milk powder normally run for about 3 months. You tend to take longer
7 with a product where the flavours are more sort of tied up together, and
8 it's really hard to break them apart and treat them individually. Things like
9 yoghurt. The flavour is reasonably simple, it's not that complicated, you can
0 manipulate the product easily to make standards, it's pretty simple.
1 Something like chocolate, we've got what we've got and if we want to try
2 something out it might be, you know, by the time it has been manufactured
3 it takes 3 weeks before people taste it to for it to set up properly and it

4 might be 3 weeks before there's time available to make that sample. So
5 any product that's liquid we can prepare here, and that really helps the
6 training process because you can listen to what the panel's stating one day
7 and say tomorrow we'll try x, y and z, and do it. But some of the things like
8 chocolate and baked products it might be several weeks before we get
9 those samples and we've got to go back to what they were talking about
0 2 weeks ago. The chances are they will have changed their minds anyway,
1 it's a much more difficult process.

2 KE

3 How much variation do you think there is between the group leaders in
4 terms of when they declare a panel trained?

5 Expert 2

6 I don't know. In theory, we should stick strictly to the standard deviation
7 rules and things like that and we should all be doing it the same. Even
8 though that's what we are supposed to do a lot of times I think it's more
9 done on gut feel and looking at the data rather than actually physically
0 sitting down and doing it, you sort of get an impression as you're going
1 along and looking at their scores on the board.

2 KE

3 And that's the data you're talking about, the gut feel looking at their data
4 rather than the historical data?

5 Expert 2

6 Oh yeah, looking at their data. I would tend to say that there probably is
7 some difference but I'm not sure what they are.

3 KE

9 Mmm, or if they are significant?

0 Expert 2

1 Yeah.

2 KE

3 So leading on from that, how do you feel about this process?

4 Expert 2

5 Which process?

5 KE

7 The one that we're going through at the moment trying to document...

3 Expert 2

9 It's quite interesting actually. I mean I'm quite sure that we all have
0 different styles in taking a panel, there's not much question about that. So
1 it is quite interesting some of the things that you have asked and that you

2 talked about at the meeting whenever it was, a week or so ago. It made
3 me think, do we do things just because it's historical. Why do we do them.

4 KE

5 Mmm yes, well when you have an outsider come in, they are allowed to
6 ask dumb questions.

7 Expert 2

8 Yeah that's right and it does make you stop and think, you're right, why do
9 we do that. Yeah, it's interesting.

0 KE

1 So following on from the question, is there variation, I take it there is no
2 feedback system to you from central within the unit on performance of
3 individual panels that you've trained compared to other people's panels.

4 Expert 2

5 There is in a sense. It's a difficult question to answer really. There are
6 what we call research meetings about once a fortnight and each of us talk
7 about the problems we're having and the progress that the panels are
8 making or not making as the case may be. If anyone's got any suggestions
9 then they make them and that sort of thing. So then you sort of get to hear
0 and evaluate for yourself I guess, like when the chocolate panel was having
1 problems I was having a moan about it every meeting for quite a few
2 weeks and the others all made comments or suggestions or said, yes, I
3 know exactly how you feel, I've got similar problems with such and such.
4 So there is that. I guess you can't necessarily compare one panel directly
5 to another because they might be trying to achieve quite different
6 objectives. The other time that it comes out is at our performance review,
7 like if I took a whole year say to train a chocolate panel then that would
8 come out in my performance review. Okay while I might have been doing
9 everything that I could think of it probably would be that I should have
0 been seeking more help or getting more support or... admitting defeat. I
1 don't know, because to achieve at least half of the objectives that we have
2 your panels need to be performing and if they're not then those objectives
3 will suffer and that comes back in that form.

4 KE

5 As a general comment then, would it be fair to say that the feedback that
6 you get is more qualitative than quantitative?

7 Expert 2

8 Within the section yes. We can get quantitative feedback from the data but
9 that is up to the individual. I mean if I want to know how a panel is doing
0 it's my responsibility to go and look at the data and be honest with myself,
1 so yes, mostly qualitative.

2 KE

3 So just getting back to the training process. I know that you mentioned 3
4 sessions. What is the normal duration of a session and how many sessions
5 a week.

6 Expert 2

7 OK, normally 3 sessions a week and they last anywhere from... they can
8 last up to an hour. I find them pretty hard going when that happens, so
9 about 45 minutes is my tolerance although it does depend on the group.
0 If we've got a big group they take longer and like, for a panel that is
1 trained like the butter panel they still have about one training session a
2 week or a fortnight, depending on their schedule. Some of them are quite
3 quick, 15 or 20 minutes because there isn't that much... it's more for them
4 so they can see how they score compared to everyone else. I would say
5 that the average is about 40 minutes.

6 KE

7 What effect would it have on their performance if you varied the duration
8 or frequency of their training?

9 Expert 2

0 I don't know. The main reason as far as I can see that we have them on
1 regular times and days and things is so that they get into a routine. In the
2 institute, when we had staff as panellists we just used to ring them up and
3 say can you come today not tomorrow and that was fine, that was great, it
4 had a lot more flexibility but because...I mean you can actually never
5 contact people by phone when you want to here and so as a regular thing,
6 people know where they stand and they know that they're expected to turn
7 up.

8 KE

9 I guess my question is, why is it 3 sessions a week and not 4 or 5 or 2

0 Expert 2

1 OK. Again partly historic. 3 sessions... the butter panel sometimes does 4
2 sessions because of the high workload that they have. We decided that if
3 you only did 2 sessions a week you didn't actually achieve that much
4 because you might start of on say Tuesday working on caramel flavour and
5 it might take 3 sessions or more to work through that. If you only have 2
6 then a big long gap, you will have to repeat what you did on the last day.
7 4 we decided was asking a bit too much of people. we didn't there was
8 enough dedication in the community to get that, so yeah, 3 was a trial thing
9 but it works really well. Trying to achieve a balance between making it
0 worthwhile for people to come but not being too big a tax on their time.

1 KE

2 Talking about making it worthwhile for them to come, how do you pay
3 them?

4 Expert 2

5 We pay them \$10 per session regardless of the length of the session so we
6 do 10 minutes or an hour, they get \$10. They get paid by cheque
7 fortnightly. Well technically they're not paid, it's reimbursement so they
8 don't have to pay tax on it. Legally they're volunteers. Also with some
9 products we bribe them with food. Like the butter panel might get
0 chocolate afterward because when they first started they all moaned
1 terribly and were really offended that they would have to eat butter and
2 other people got to eat cake and chocolate, which were the 2 other panels
3 set up at the time and they really thought that they'd got the short straw.
4 Nicole's panel which she and Expert 4 are doing at the moment is a really
5 awful product, it's assessing bitterness in protein solutions and it's
6 disgusting so they need a bribe factor. I mean it just tastes so unpleasant
7 that I don't actually think that money makes up for that. And the butter
8 panel have been going a year now and so we got them all a book voucher
9 and just little small things.

0 KE

1 Right, so at what point do they start getting paid?

2 Expert 2

3 They get paid right from the time they come to screening. They get paid
4 less for coming to screening. Normally if they came to a 5 day screening
5 they get \$30. And they get paid right through their training, that sort of
6 thing. We're also starting to introduce some sort of bonus scheme, not on
7 a regular basis but just every few months for those people that have
8 attended more than say 90% then they'll get a small bonus like a book
9 voucher or a petrol voucher or something, just a little extra.

0 KE

1 It's an expensive business though

2 Expert 2

3 It is, really expensive

4 KE

5 Presumably, that's all passed on to the client?

6 Expert 2

7 Mmm, although I'm not completely sure on the workings of it but my
8 understanding is that for the initial setup of a panel, you can have multiple
9 clients and the institute in fact will be paying for that, that'll be as part of
0 the working expenses of the unit but then if we go training for a specific
1 project then that's project work.

2 KE

3 How many panels a year do you train? Say you personally.

2 being... sort of like doing a quality assurance type plot. Very hard to
3 identify, particularly because like we might do it one day and they'll be OK
4 for a while, so you sort of have to allow for that flexibility. When we first
5 set up the unit we sort of said right, a nice fresh slate, anyone that doesn't
6 perform is going to be out the door sort of thing, but it isn't that straight
7 forward 'cos quite a lot of it has got to do with personalities and things.
8 I find it difficult because if I think someone is trying then I should keep
9 giving them that opportunity to try and understand, that sort of thing. I
0 find that quite hard to deal with. Expert 3 has had to remove someone
1 from one of her panels and I'm going to have to do it with chocolate,
2 there's just no question of it.

3 KE

4 What is it that will finally cause you to remove a panellist. You personally?

5 Expert 2

6 For me it is that, well in this case, we've done various projects for people
7 and the results have been confounded by the variability of the panel and
8 it looks like there's something there but there's so much variability in the
9 data that you can't draw that conclusion confidently, and so you've got a
0 client saying, look, we're paying you x thousand dollars to do this and you
1 can't tell me because your panel is so variable that you're not sure. And
2 that reflects badly on me and on the section and on sensory as a whole and
3 so to improve that variation within the panel has to be dealt with.

4 KE

5 Before the results go to the client, presumably.

5 Expert 2

7 Yeah, well in the future. In this case we've told the client that we're just
3 not doing any more project work until we fix it and they're pretty unhappy
9 about that but you know I had to say to them, look, either you want to
0 keep getting data that you're not going to be very confident in or you're
1 going to have to have a break for a few months.

2 KE

3 Are you the only people who do this sort of work around?

4 Expert 2

5 There are a few other places.

5 KE

7 In Manawatu?

3 Expert 2

9 Establishment 2 do a little bit. They don't tend to do fully trained panels
0 as much, to my knowledge anyway.

491 KE
492 When you've got a person that's performing well, you trained them up and
493 then performance starts to decline, what causes it?

494 Expert 2
495 Probably depending on the individual. Some things are that they get bored,
496 on a tasting, they sit down, write down a few numbers, take the money and
497 go home again sort of attitude. Maybe not enough on-going training like
498 if I put them in the booths week after week and don't give them any
499 feedback. People may start going off on their own tangent. Stress,
500 pregnancy... got one of those at the moment.

501 KE
502 A stressed out pregnant person?

503 Expert 2
504 I don't know whether she's stressed out but she's pregnant and it's just
505 amazing... absolutely complete turn around in what she's got

506 KE
507 What will you do with her?

508 Expert 2
509 Well she actually came and said to me that she thought she wasn't doing
510 very well about the same sort of time that I started noticing and I said, well
511 we'll see how it goes for a few more weeks because she ad been having
512 really bad morning sickness and she thought she was over that and it might
513 stop. I don't actually think it has though so I'm going to suggest to her that
514 she comes back and get some re-training after she's had the baby or
515 something like that.

516 KE
517 Is that the normal thing to do with a pregnant person?

518 Expert 2
519 I don't know because I haven't had many of them. But she's just getting
520 totally different results to the rest of the group. There's also personal
521 health. Every panellist has their bad days or bad day or two. Yeah, my
522 personal point of view is that it has to do with their state of mind when
523 they come in. If they rush in here in a hurry to get somewhere to pick up
524 the kids and the other kids are all fighting, and they just sit down, race
525 through it and go again, then we'll expect that we're not going to get as
526 good data from them as if they're in a calm, relaxed atmosphere. Some of
527 the panellists have actually commented that... see often a lot of the
528 samples are really similar... and after maybe the second week of it they sort
529 of start to think, gosh there must be something wrong with me, these all
530 taste the same, you know, I must be missing something, and they start
531 getting really paranoid I guess.

532 KE
533 Yeah, this is where feedback becomes important obviously, to tell them
534 they're doing OK.

535 Expert 2
536 Yeah

537 KE
538 What do you do with people who are good performers but through fatigue
539 or adaptation they're just losing their sensitivity to a certain product?

540 Expert 2
541 I guess they're probably quite difficult to identify in some respects but if
542 you are having regular training sessions, it might come through then. Also,
543 like making sure that they're remembering what those reference samples
544 are and they're scoring things according to those samples. Panellists have
545 a real tendency to want to score samples relative to one another on the
546 day and so therefore you can never get day to day consistency. So yeah,
547 going back over those reference samples regularly, I think can help stop
548 people becoming desensitised but you also get the opposite effect that they
549 can get over sensitised and tell you that they can taste a hundred and one
550 different flavours in the sample and maybe they are or aren't but no-one
551 else can taste them or you know... yeah, I think reference samples are
552 about one of the best things we can do.

553 KE
554 What about for a training panel, what's the most work they could do in say
555 a week. Do you put a limit on it?

556 Expert 2
557 Yeah, we still limit it to the three or maybe four sessions per week and for
558 most products we limit it to about four samples per session.

559 KE
560 What I'd like to do is actually move on specifically to your tasks as the
561 group leader. What I'm trying to do, perhaps I can just show you the
562 methodology that we're working through. This is part of the idea that I'm
563 working through. basically the stage that we're in at the moment is the
564 phase 1 box where we are looking at developing a set of tasks and task
565 descriptions, well hopefully the task titles are self explanatory and then the
566 items table, and the items are the key descriptors for each of the tasks. It's
567 not unlike what you do with a panel when you're trying to agree on what's
568 bitter. That's... we don't have any particular agenda in any one session that
569 we spend together, we just go as far as we go and stop, and carry on from
570 there. So what I'd like you to do is just ask you to think about your role as
571 the group leader. Where does it start and where does it finish?

72 Expert 2

573 I guess it starts when we have enough requests programmed waiting for a
574 certain product type to justify setting up a panel and the decision is made
575 to do that and someone has to be responsible for it. That's where it starts.
576 Where does it finish. I guess that the end of the line is that you're
577 responsible for producing that data or having the data produced and
578 reporting it and being accountable for it, in that you can understand,
579 explain and justify what was said about it. To me that is the end of the line
580 for that panel. Once you've done it I don't think there ever is an end to it.
581 People will ask you about things you did 5 years ago, you know that you're
582 no longer responsible for but they still want to know and pick your memory
583 and that sort of thing on whatever information there is.

584 KE

585 Right. Because it's so open ended what I'd like to do to make this project
586 more manageable and give you something that will be usable is to actually
587 tie it down to something pretty tight and so the focus will be on the
588 training of a panel rather than the operation of a fully trained panel. Can
589 we agree on an end point perhaps which is really one side of the domain
590 of knowledge which is say where you make the decision that they're fully
591 trained. Is that a fair... you know, the edge of the cliff? and say, on the
592 other side is operational panel, this side is training panel.

593 Expert 2

594 OK, that's fine.

595 KE

596 Coming back the other way, at the start of the process, you've got the
597 clients coming in or products to evaluate, the decision is made to set up a
598 panel, your active role wants your committed to the panel begins there
599 with the advertising and the screening. Is that...?

500 Expert 2

501 If you want to tie this down to the training process, then it would either
502 start there or maybe at the screening time. I mean, there is stuff to do
503 before that like finding out about the projects, what kind of tests I'm going
504 to need or that sort of thing. That is pretty much project and product
505 dependent so yeah, if there is time we can screen panellists out here. I
506 mean panel leaders weren't responsible for advertising or organising panels
507 cos they were all one huge group. But this time round we have been a lot
508 more involved in organising ourselves.

509 KE

510 So is the screening generic or is it related to the product?

511 Expert 2

512 It's a mixture. Like a generic base but with things that can be added on or
513 taken off depending on what it is we want to achieve.

614 KE
615 OK so it is fair to say that the way that you run the selection is product
616 dependent to a certain degree? The preliminary project work does have
617 something to do with how you set up to screen? How about we, just for the
618 time being, we can change it, we just say that the start point is the project
619 plan and goes forward to the point where you say they are trained.

620 Expert 2
621

All right

622 KE

623 So, tasks that you do, it seems to me that the first one is a planning task.
624 is that a fair comment?

625 Expert 2

626 Yeah although I would say that in some instances we just know that say
627 there's going to be heaps of cheese work and we're going to set up a panel
628 and the project information comes afterward or during the training process
629 and other cases we know before we start like Nicole and Expert 4 are the
630 classic examples. They know exactly what it is that the panel is required to
631 do, OK it's for 1 specific project OK, but other panels are set up to deal
632 with multiple projects so the point that the project planning comes in does
633 vary. Hopefully we normally know something about the project we're going
634 to be dealing with right from the start.

535 KE

536 OK, well I'll just note planning and write that it's variable. Following on
537 from the planning you get a huge response to the ad, hundreds of people
538 turn up. What do you do?

539 Expert 2

540 Well probably in response to the ad, you have to sort out times of the day
541 when you're going to run your screening, how many people you want to
542 attend, how long it's going to take them, the equipment that;s going to be
543 required and the samples and everything required. Any publicity
544 information you want to give them. Who's going to do what tasks on the
545 day. All that sort of nightmarish stuff. During screening we have 2
546 approaches. One is that you can plan it all out and say right this is what
547 we're doing on days 1 through to 5 end of story. Or if you're about to do
548 something that you don't have much experience in, what we might do is say
549 plan the first two days and go through the results and say well the
550 screening is far too hard. No-one can pick the differences between
551 examples, it's too difficult. By making it too difficult you're not going to
552 achieve anything, because you're not going to select anyone. So we'd
553 normally want to try and find a level which would spread people out and
554 show their different abilities to discriminate Ok. So you've got to try and
555 find that level fairly quickly, because we only have 5 days to do it. We have
556 a lot of base knowledge with which to do that, so it's setting up the actual

657 screening...

658 KE
659 You mentioned earlier on that you interviewed you potential panellists.
660 Who does that?

661 Expert 2
662 The panel leader does.

663 KE
664 That's a huge task then

665 Expert 2
666 Yeah, it is quite a huge task. One of the reasons we do it was in the first
667 lot of screening we did here, we didn't and there were a whole lot of
668 confounding reasons but Nicole had been responsible basically for co-
669 ordinating all the people factors. She knew most of the names and who the
670 people were and blah, blah, blah and then the rest of us came and sat
671 down and tried to choose people and we were looking at their score and
672 thinking, you know, specially if they were on the borderline, and you think
673 well is there anything that's going to swing this person one way or another.
674 Nicole was having to try and describe people to us and remember them
675 and we decided that that was really a bit flawed. I mean they were only 10
676 minutes, just to try and... each person to develop an impression of...

677 KE
678 Could I work with this person?

679 Expert 2
680 Yeah, but also before then I had a look at their data and sort of pencil in
681 those people that I thought had potential and so trying to take particular
682 note of them and to see what they'd be like. There was one person I chose
683 whose screen was quite good but she has quite a few clashes with her
684 Massey timetable and she's also doing public relations and her personality
685 is just really good for the group dynamics, she really listens and she is
686 responsive and asks good questions and you know, and I decided that she'd
687 be worth having even if she drops out, just to get the group going. She can
688 only come twice a week instead of 3 times a week but it was a trade-off
689 which I decided to make based on her personality that could contribute.

90 KE
91 So as a task, the group leader sounds like an interviewer, but also a
92 selector. Are they appropriate titles or should they be lumped together?

93 Expert 2
94 Interviewing is an aspect of selection.

95 KE

696 OK, so could we list a task as selector? Personnel selector?

697 Expert 2
698 Mmm

699 KE
700 OK, moving on down, what other generic tasks do you do...

701 Expert 2
702 ...While we're screening or after?

703 KE
704 Just moving on through, after you've done the task of planning, and done
705 the task of being personnel selector and I guess that goes on because
706 you're still working out the out-liers right the way through. What happens
707 next, what's the main task?

708 Expert 2
709 A couple of things I'm thinking of really, I'm not quite sure which order
710 they come. One is that of group co-ordinator. Actually standing up there
711 and trying to see the wood for the trees. The other one is more of a
712 physical task I guess like organising the samples, co-ordinating them, trying
713 to select samples which might help the group in what we're trying to say.
714 At the same time reading the literature or past reports or whatever and
715 trying to understand that from a food science point of view but also trying
716 to be a facilitator, you know, without trying to put the answers into their
717 head.

718 KE
719 Right, so facilitator seems to be one of the roles, getting things happening,
720 group leading, controlling the discussion and keeping things going.
721 Throughout this co-ordination or administration role, how could we best
722 describe that as a task?

723 Expert 2
724 I'm trying to think of a good word for it. One part of it is the researching
725 side, trying to understand any reactions that might be going on in the
726 product, reacting compounds and that sort of thing. The other one is
727 maybe a little bit more on the technical side and may actually be delegated
728 to some one else, in that it's the physical handling of samples, checking out
729 sample availability, preparing sheets fro training because often during
730 training we might go through half a dozen sheets or more.

731 KE
732 Could we reasonably say that that is part of your planning task? The
733 planning goes on throughout whether you do it or you get someone else to
734 do it, it's your task to plan what samples come out. Researcher sounds like
735 something different though, it seems to sound like something that has to

736 be done...

737 Expert 2
738 Yeah, it is different.

739 KE
740 OK, well let's note that down. Good.

741 Expert 2
742 There's the analysis and interpretation of the data that's being generated
743 during training. In some ways you could say that's part of planning, in other
744 ways you could say it's part of personnel selection...

745 KE
746 Right, and then it has the function that's just related to the client and
747 product as well, you've got to come up with good results.

748 Expert 2
749 That's right, so in some ways it's sort of like a separate thing.

750 KE
751 Right. They're all interrelated because otherwise it wouldn't be one
752 person's job. Could we say that analyst is a task?

753 Expert 2
754 Yep. What else do we do. General communication, like communicating
755 with Expert 1, communicating with clients, communicating to the other
756 staff here as required. Obviously in relation to the panel that's facilitation
757 but just keeping the whole thing tied together really.

758 KE
759 Mmm. That sounds like, what we would call in computer terms, the project
760 manager, a person that makes it all happen, who writes all the reports,
761 talks to everyone, co-ordinates... is that a good title for a task?

762 Expert 2
763 It is, I know exactly what you mean but in institute terms it's not the
764 project manager, it's usually a different term. I would basically could it just
765 communication, I mean it is project management, yeah that would be OK

766 KE
767 Well we can always change it as we go along, I mean this just records our
768 thoughts. Any others? You're thinking, he can't possibly want any more...

769 Expert 2
770 Let me think. Not that come to mind.

771 KE

772 OK, well the nice thing about this is that in developing a knowledge base,
773 where they differ from traditional databases is that they grow and develop.
774 Well that's good and I think that we've probably gone far enough for 1
775 session. Taxing isn't it trying to answer what you know inside but are not
776 usually asked to surface all the time, you just do it. Perhaps what I can get
777 you to think about Expert 2, between now and our next session is the
778 facets of each of these tasks...

779 Expert 2
780 What makes them up?

781 KE
782 Yeah, what are the elements that make up interviewing. An element might
783 be something like social skills. If you are a sort of nasty, aggressive person,
784 you can't you can't be an effective interviewer so perhaps that's one. If you
785 could think through the tasks that you do and try and isolate elements you
786 might find the same elements occur in several tasks like communication.
787 So I'll see you next time and thank you.

788 Expert 2
789 Thanks.

790 **DURATION 58 MINUTES**

791 Expert 2 - SESSION 2

792 KE

793 OK, if you could begin by telling me in general terms what you do in those
794 last few stages immediately prior to releasing the panel for project work.

795 Expert 2

796 Well, several things I guess. One is, which I reckon is really important is
797 giving the panel confidence and showing them that they are doing well and
798 that their scores are meeting whatever criteria we've set, whether it's
799 standard deviation or range or whatever we've chosen. We try and do that
800 by giving them feedback sort of like that day or the next day on how well
801 they did and if we repeat that part, show them what their scores were over
802 certain days or whatever, just so they feel good about themselves. Other
803 things are just to be confident in myself that each of the panellists do know
804 what the attributes are and that they can score them reasonably well and
805 that should have been achieved by the use of standard samples and things.
806 But maybe repeating ones that they hadn't done for a while or something
807 like that. Another thing that I would do is possibly start bringing in project
808 samples so if cake is going to be something unfamiliar which they haven't
809 come across before, which quite often happens, or it's like what they've
810 had before but because it's not a standard sample which we've
811 manipulated, it's not a naturally occurring flavour or whatever. It might be
812 in the same flavour group but not exactly what they're used to. They have
813 to learn to recognise that and know where to put it, rather than sticking
814 everything in the 'other' category because it's not exactly as they've learned
815 it, so that's quite important. Maybe telling them about upcoming projects
816 and what sort of feedback they're likely to get. We don't normally tell them
817 too much about the aim of the work or anything like that because it can
818 result in quite a lot of bias if you do that too much. It's difficult but it's
819 good to try and find some sort of balance so they know a little bit about
820 what they're doing but not so much that they then skew the data for you.
821 Those are probably the main things.

822 KE

823 That's good. In relation to you actually confirming whether they understand
824 the attributes and can scale them or not, what actual range of tests do you
825 put them through to do this?

826 Expert 2

827 Traditionally, it's just a round table training session like in here and yeah,
828 giving them reference samples and making sure people can recognise
829 them... like in training they might be doing say a flavour which we call
830 'cow-ey' and they might have had it at three different levels, OK and they
831 get those often in a similar order and if you suddenly throw one in amongst
832 lots of other samples they still should be able to recognise that. So it's
833 doing things that they're not used to so much like often in training you
834 have a control sample first as a reference point for them but when you're

835 doing project work that doesn't happen, well not all the time and so
836 suddenly taking that sample away what happens to their scores? The other
837 thing to do is to put them in the booths and make them do it by
838 themselves because a panel that does all their work in here for 3 months
839 and all of a sudden they have to do it without any visual feedback, they
840 often just fall apart for a couple of days. It's amazing because they think
841 that they're not getting any feedback or support from the people around
842 them, but they must be because, yeah, they just fall apart. Even if it's just
843 they might have scored the sample and then look at their neighbour's sheet
844 to see, it doesn't mean that they're cheating, that they're going to change
845 but they just get that supported feeling, that they can look and note how
846 far off they think they are. What would they do in the booths... Other
847 things we could do involve... oh no, I was going to say ranking tests but
848 that's more of a training process.

349 KE

350 So, when you are assessing their ability to spot and rate an attribute how
351 many exposures would you give them with each attribute before you say
352 that the result is consistent?

353 Expert 2

354 You mean at the end of the training process?

355 KE

356 Mmm

357 Expert 2

358 A lot of it would be based on what I thought had happened throughout the
359 whole training process but for me, in every panel, there's always at least one
360 attribute which I'm a bit uncomfortable with cos I think the panel
361 flounders a bit. So I might give them more leniency on that, but like for
362 butter and saltiness, I'd expect them to be really consistent and really good
363 and so at the end of the training, I'd expect only to have to do it a few
364 times because their training should have been good enough, but I have
365 confidence in that and it's just a final check. It isn't something that we do
366 that formally, just the last few sessions we take particular notice and maybe
367 give them some of their data back and that sort of thing. Point out to the
368 people if they're consistently scoring lower than everyone else or whatever.
369 In terms of the number of times... I don't know... say 3.

370 KE

371 You feel that whether or not you're doing it, 3 would be enough exposures
372 on an attribute to see whether they were getting it right or not?

373 Expert 2

374 It really depends on the circumstance. Because it has been a built up thing
375 over 3 months, then it's only a final check, but yeah, I'd probably want to
376 spend 1 or 2 weeks say, before they started project work, checking all that

377 kind of thing, maybe actually getting some project samples but not starting
378 the project yet, and checking all the data. Yeah, that would be 6 sessions
379 and you can check on a number of attributes within a session obviously.
380 Yeah, probably 3 -ish, anywhere between 1 and 5 I guess.

381 KE

382 Sessions?

383 Expert 2

384 No, times that you'd look at an attribute.

385 KE

386 And how many attributes would you look at in a session?

387 Expert 2

388 At the end of the training, I'd be going through all of them unless I'd done
389 something really odd like thrown in some margarine or something.

390 KE

391 Is this for a butter panel?

392 Expert 2

393 Yeah.OK because at the end of it, they should be able to focus on every
394 attribute in the sample, saying is it there, isn't it, how much. Yeah in the
395 last week or two I'd be looking at every attribute, each time but I might
396 be... I'm not being very clear here. When I looked at the data, I'd look at
397 every attribute, but the sample that I gave them might have focussed on
398 one or two, like I might have given them two butters with different
399 saltiness levels and two that had different degrees of oxidation. OK
400 something like that. there's only so much you can vary in one session.
401 Looking at the data, I'd see if varying the level of oxidation, did their
402 saltiness scores go out the window? or did their previous scores?

403 KE

404 What is an acceptable result for you?

405 Expert 2

406 It depends on the panel, and the attribute and the product. Normally we
407 like to have the standard deviation round plus or minus 1, that's on a 10
408 point scale. But like if we have an attribute which we know is the weaker
409 points of the panel then we can allow but 1.5 is pretty much the end of the
410 line. Getting above that makes your data pretty fuzzy.

411 KE

412 Mmm. Why do you use a 10 point scale?

413 Expert 2

414 Tradition. I mean there's a reasonable amount of research looking at

015 different sensory scales and by the time you take into account avoidance
016 of end of scale, central tendency and that sort of thing, some people prefer
017 using 15 point scales or using a category scale. Some researchers. because
018 they think it gives the panellists more freedom, more flexibility, and if you
019 have a wider range you're more able to discriminate between samples. But
020 other researchers use a 5 point scale. We have had panels that used a 5
021 point scale and have found that it's really hard to get discrimination
022 because everything is scored 2,3 or 4, so it makes it really difficult. Another
023 scale that we use a bit is a line scale, which isn't a category scale, it's a
024 continuum, and that's 150mm long. The standard deviations are a little
025 different on that.

026 KE
027 Is it valid to put the data together off a line scale and a 10 point scale?

028 Expert 2
029 No.

030 KE
031 Mmm. Who decides what scales are used for a particular panel?

032 Expert 2
033 Historically, the section has always used category scales and why, I don't
034 especially know. Probably because they are the most common in sensory.
035 Line scales have been used, well I know why a line scale was chosen for
036 chocolate, and that was because that's what Cadbury's do and if we wanted
037 to do work with their laboratory in the future then we'd be compatible.
038 We've used a line scale for testing things like testing the spreadability of
039 butter because the panel at the time felt that, like spreading was more
040 important than putting it into categories. It was more of a relational kind
041 of thing and so a line scale was more appropriate. But really what kind of
042 scale you choose depends on what sort of project you're going to be doing.
043 I mean like Expert 4 with the bitterness panel has chosen to use a
044 technique called magnitude estimation, I don't know if he explained that
045 to you.

046 KE
047 He has.

048 Expert 2
049 He has, yeah, and that's because of the kind of work he's going to be
050 doing. the range is so large it would be difficult to fit it into a 10 point
051 scale, so really the scale should be determined by that. Difficult if you're
052 training say a butter panel but the project is going to require very different
053 grading systems. Then we may teach them how to use different scales and
054 that's quite hard.

055 KE

156 Could we just go back to the testing phase, the bit at the end of the
157 training cycle. What happens if you have the gut feel that this is the right
158 time for them to finish, or there's a bit of commercial pressure from a
159 project sitting there, and the performance simply isn't there, like the
160 standard deviations are beyond 1.5.

161 Expert 2

162 Sometimes the work goes ahead anyway but normally we tell the person,
163 look, we can do it but the panel, you know, isn't up to scratch but if that's
164 what you really want to pay for then, you know... but normally, no, it hasn't
165 happened that many times. Normally, we would like... if it was a serious
166 problem, like if they had standard deviations of 3 or 4, then you would
167 have to look at are the reference samples inappropriate or maybe you've
168 got some panellists that just aren't up to the task. I mean, when it gets to
169 a situation like that, you really have to start looking hard at individuals. Is
170 the standard deviation high because everyone is putting 2 or 3 and one
171 person's putting 8. Or is it because the panel is completely spread over a
172 wider range. And if it's 1 or 2 individuals, yeah that's quite hard. You need
173 to think, well did they have a bad day? Have they been slowly going off
174 form for a few weeks. And repeat it, are they doing it again. If they are
175 consistently causing the performance of the panel to be low, then they
176 need to be told that, and if it's something that they can't help then they
177 either have to learn to adjust or else they need to be off the panel.

178 KE

179 Right. So how long would you persevere with them before you chopped
180 them off the panel?

181 Expert 2

182 Too long. Yeah, I've got a couple like them at the moment and its'... for
183 me it's a really difficult thing. When we use to be out at Establishment 1
184 we used to use staff members on the panels and it was very difficult
185 personally and politically to get any people off the panels if they were
186 useless because it was seen as though they were doing you a favour and
187 you didn't have the right to kick them off and that sort of thing so what
188 tended to happen then was that you would avoid asking them and if they
189 got the hint then they'd avoid coming and then slowing you'd just sort of
190 drift apart. When we came here we said OK, if we have a useless panellist
191 we're going to get rid of them but so far Expert 3 is the only one who has
192 actually done it, I'm supposed to be doing it this week. I think because we
193 invest a lot of money and time in people, it is worth persevering with them
194 for quite a while, even if it's really difficult and you are in that situation
195 when the project's really urgent and there's one person. Maybe they can
196 keep coming and you can disregard their data until you try and sort it out.
197 I guess that after... like if you've been telling them quite clearly and
198 consistently and if after 2 or 3 weeks of training, then they still have that
199 problem, au revoir. But yeah, it's something that I find really difficult to
200 have hard and fast rules for...

001 KE
002 Presumably a life change situation that goes on beyond a couple of weeks
003 is probably going to go on for quite some time like pregnancy or...

004 Expert 2
005 Yes, that's a classic around here...

006 KE
007 So do you keep individual records of the panellists to track their
008 performance?

009 Expert 2
010 No, we don't. Sometimes, like for a specific project, I might do an analysis
011 and look at individuals but it's not something that we do standardly
012 although I hope that when compusense is installed, we will be able to do
013 that a lot more easily. At the moment it's really time intensive so you
014 know, your own things always take a lower priority which they shouldn't
015 because that's a very important thing because if you can show why a
016 particular panellist has been gradually going off on their own tangent, then
017 you can stop it much more quickly, but if you don't notice it till after a
018 couple of months then it's often too late.

019 KE
020 Mmm. Well that covers the case of an individual who is causing variation
021 within the group. What about if there's an attribute which is causing
022 variation, what do you do about that?

023 Expert 2
024 It really depends on how you think it has been caused. The first step is to
025 do more training on it and to try and find out from the panellists if they
026 think that they really understand what that attribute is because particularly
027 here I think, there's a real tendency at the start for the panellists to want
028 to please you and so they say, you give them a sample, and you say we
029 thought the sample might have whatever characteristic and they go yes, yes,
030 yes and if you think yay, I've made a breakthrough and if they haven't been
031 completely honest with you, then you might think that they understand that
032 standard and everything's home and hosed. Particularly, well I found it
033 more here than at Establishment 1, panellists want to learn by recognizing
034 what the sample is and then by scoring it hoe they're supposed to rather
035 than recognising the flavours... do you understand what I mean?

036 KE
037 Yep, so it's going from the brain to the tongue

038 Expert 2
039 Yes, not the other way round and that causes a real problem and it has
040 caused problems with the chocolate panel because they get an unknown
041 sample like in project work, they suddenly don't know what it is and they

042 don't know how to score it because they've learned the process the wrong
043 way round. You have to ask the question, is it because somehow you've
044 tried to impose a word on them that they're not comfortable with. Is it that
045 the references aren't appropriate. Is the attribute even really relevant to
046 the product? Hopefully by that stage you wouldn't still have it if it wasn't.
047 If it was something say in chocolate like sourness, which isn't a particularly
048 important attribute because it doesn't occur very often then we might talk
049 to the scientist or client with a view to dropping it or having the panellist
050 score it under other if it occurred occasionally but not that often. If it is an
051 attribute that's really important or fundamental to the product or sort of
052 like the key to what we're trying to find out, then that's a lot more difficult.
053 Really you have to go back and look at what you used as training
054 standards, whether with hindsight something else could have been more
055 appropriate, whether in the experimental product it's maybe a sub-group
056 of that flavour that's coming out or whether it has gone beyond that
057 flavour into something new. That kind of area, trying to really just look...
058 if it's one or two individuals then it's slightly different than perhaps if the
059 panel in general is having trouble with an attribute. If it's one or two
060 individuals then you have to look at whether they have at the same rate as
061 the other people or as well or have the same understanding, are they really
062 understanding the scale or are they even understanding what's being asked
063 of them. And if not then you have to make a choice as to trying to fix it or
064 to take them off the panel

065 KE

066 Mmm. How often does that occur?

067 Expert 2

068 Taking them off the panel? Not very often. Not as often as probably it
069 should in theory in an ideal world. We have big problems with panel
070 turnover and that kind of thing and trying to maintain a stable group so
071 yeah, I have 1 chocolate panellist at the moment who's... yeah, in an ideal
072 world she would be down the road and so, I'm still making that decision.
073 I'll have to see how she goes early next year cos they've actually finished
074 now but if she doesn't improve then, there's something... like she tries
075 really hard and that's why I try to give... if I think they're really trying I try
076 and give them the benefit of that and help them to learn or make progress
077 or whatever... that's what she has got in her favour but there'll have to
078 come a point where she's becoming detrimental to what's going on. So,
079 how many people have I got rid of? only about 2 off one panel. Expert 3's
080 got rid of... sounds terrible... taken off the panel I think 1 person off the
081 baked products panel and I don't know if Expert 4 has taken any off yet.
082 I know he's looking at it. If we had, you could afford to be continuously
083 retraining people, which we virtually are, but even more so, yeah, you
084 could take people off. It also creates quite a bit of, I don't know, fear
085 amongst the others. Especially the ones that are really conscientious and
086 are paranoid that they're not doing it well enough or something like that
087 so yeah, it is quite difficult.

088 KE
089 Could the removal and the ensuing paranoia on the remaining panellists
090 transmit itself into the results?

091 Expert 2
092 Yes, for sure. Had the example yesterday where one panellist said to me
093 well, you told me I was scoring that attribute too high so now I just score
094 it on low. Every sample, she's just putting it on low. I was speechless and
095 when she said that, that I had told her, all the panellists, you know their
096 reaction was just like, has someone been talked to? you know

097 KE
098 And not me. Why not me?

099 Expert 2
100 Yeah, why her. And before when I've talked to them about I've been doing
101 lots of plots of their data so I can see where each individual is relative to
102 the mean, they ask straight, and does that mean you might get rid of some
103 of us? And when will that be? and how will we know? It's a bit sad really

104 KE
105 Yeah. When you've got people who just don't understand a particular
106 attribute no matter how hard they try, in hindsight, just what sort of things
107 cause that situation to arise?

108 Expert 2
109 A couple. One could be if they have some sort of chemical blind spot in
110 which they're just not capable of perceiving the chemical that's causing that
111 flavour. That's very difficult to determine unless we knew what's
112 responsible for most of the flavours we're testing and used those during the
113 screening. That would be the only way to really overcome that, we're not
114 technologically advanced enough to do that at present. Another quite likely
115 cause is that, for example in chocolate, we measure burnt flavour and
116 there's always been one or two people on the panel that the first time we
117 give them a burnt sample they say it's nutty, OK. Don't ask me why but it
118 has happened. And if I as a panel leader or that person fails to recognise
119 that connection, then that person will carry on scoring that incorrectly.
120 That happened with one person which I had on the chocolate panel. I kept
121 saying to her, well do you think that when you're tasting other, that's when
122 everyone else is tasting burnt? And she kept saying no.

123 KE
124 Right, so it goes right back to the initial training and what she understood
125 that term to mean

126 Expert 2
127 Yeah, and she just kept scoring this flavour under other, while everyone
128 else was scoring burnt and in the end she couldn't recognise that, couldn't

129 make that association with the other people, and so in the end, because of
130 that, her data was always totally out

131 KE

132 So that, in that example is really a panel leader error isn't it?

133 Expert 2

134 It can be, like it took me a while but I recognised it with her in the end
135 and kept saying to her, look I think that when you say nutty everyone else
136 is saying burnt so you should score that but she just wouldn't do that so
137 yeah, I know I should have been more forceful, more assertive with her to
138 say look Lisa, I think this is what's going on and I'd like you to try scoring
139 it there and see what happens and, yeah, I don't know if I was quite that
140 assertive about it

141 KE

142 Mmm. Or just at the start if you let them slip through and then of course
143 the need for internal consistency arises, if you've said nutty for the last 3
144 months you have to kept on saying nutty otherwise you were wrong for the
145 last 3 months

146 Expert 2

147 Well I think that's probably another thing as well. Also a lot of that comes
148 down to personality I think, of the panel leader and of the panellists. Like
149 some people are quite happy to say, oh OK, I think it tastes like nuts but
150 if the rest of you think it's burnt, I'll put it there. Other people are quite
151 determined it's not burnt, they know what burnt is and that's not it. I
152 think... I'm just trying to remember your original question... it is, a lot of
153 it's panel leader error or interpretation, and yeah, I would say that one of
154 the most likely causes is through the training.

155 KE

156 How could that potential for error be eliminated?

157 Expert 2

158 The panel leader has to be really tuned in to that kind of thing in that if
159 you have a panellist who is consistently not perceiving something, instead
160 of just writing him off as useless, trying to see, is there something else
161 that's going on. I mean another classic is that people get caramel and burnt
162 confused OK, and so they might be scoring it high caramel and low burnt
163 and half the panel is the other way round. So you have to learn to look for
164 those things and say, OK are they mis-identifying that, do they not
165 understand or can they simply not perceive it. I don't... there are a few
166 compounds which we have that a few people really have problems with
167 but, in terms of pure compounds, I don't know of many panellists that
168 we've got that have blind spots

169 KE

170 When you're getting toward the end of the training phase and you're
171 starting to think these people might be ready to actually move on to
172 project work, what cross-checks are there, apart from checking the data
173 and the standard deviations, on the panel leader's performance?

174 Expert 2

175 I'm not sure to be honest. I can't think of any formal checks that we have,
176 on how the panel leader has done other than how it's reflected in the data
177 and group. And that doesn't mean to say that if the data isn't that good
178 that the panel leader has necessarily done a bad job. I think I said in one
179 of the earlier interviews that if it's been going on for a long time and
180 they're not making any progress, then that person has a responsibility then
181 to ask for more help or get more support from within the section. We had
182 this problem with... the chocolate panel has changed quite a bit since we
183 started doing these interviews and now I ask Expert 1 and Technician 1 to
184 sit in on the sessions and they sort of act as dummy panellists but it's really
185 good because they notice things that are going on and can talk to me on
186 a more technical level but from a panellists viewpoint so yeah it can be
187 really helpful and also it means that if I'm away, they can act in my place
188 and be very aware of the group dynamics and what has been said in the
189 past and...

190 KE

191 Mmm. Is that general procedure that's being followed or is that just for the
192 chocolate panel?

193 Expert 2

194 Well most panels we have tried to have like the panel leader and someone
195 else in the section who knows a reasonable about what's going on with that
196 panel

197 KE

198 Like sitting in all the time?

199 Expert 2

200 Certainly sitting in while that panel is being trained, that sort of thing,
201 because we found... well it happened with the chocolate panel because
202 Christine did it and then left and then I got them and didn't have the
203 foggiest idea what they were talking about, how the group worked, that
204 sort of thing and so we found like, when you go away or you're sick or
205 whatever, that things go much more smoothly and the panel's much
206 happier if they have someone that they're familiar with taking them and
207 who understands what they're on about and that sort of thing, so yeah,
208 we're trying to do it for most products. It hasn't happened as formally as
209 it has for chocolate with the other panels so we've certainly made some
210 effort toward that end and I think it's really worth the time.

211 KE

212 Yeah, that's a good idea. What about the situation where you start a
213 project and performance drops? You've actually said they're trained and
214 then after a while you think, oh, we can't give this data to the client

215 Expert 2

216 I would say that a panel is never perfect and I don't think I'd say that
217 they're never trained but I'd say there's always room for improvement so
218 the work I'm doing at the moment, I'm making sure that their data is being
219 collated every day after the panel and I'm putting the mean and standard
220 deviations on a graph out for the next panel so if I need to I can repeat
221 that day's tasting or I can talk to individuals who may be having a problem.
222 or if the whole panel has gone off the edge then we can stop it, have a few
223 day's training, talk about the problems and go back and start again. When
224 it happens during project, the problem arises if you're not on top of it and
225 that's something that we haven't done that well in the past but we're
226 starting to be a lot better at it because if you let the data mount up for
227 three weeks while you're doing project work and then find that half the
228 days are absolute garbage, I mean OK you could go back and start again
229 but a lot more extra work and a lot more expense so it's, yeah, you're
230 much more in tune with what's going on if you're looking at it daily.

31 KE

32 So, timely monitoring of the data seems to be essential?

33 Expert 2

34 Yeah, it is

35 KE

36 How has the installation of the computers helped?

37 Expert 2

38 I can only really talk for my panels. I think that, say that for Expert 3,
39 they're helping quite a bit. The butter panel is going through a retraining
40 phase so they're not doing that much work on compusense and with the
41 chocolate panel it hopefully will be useful but it's not proving to be ideal,
42 for a number of reasons, I don't know if you want me to describe them

43 KE

44 Sure

45 Expert 2

46 One is that compusense measure it in pixels and it's limited to 61 pixels,
47 whereas before we had a 150 mm line, so we have reduced the sensitivity
48 of the scale by more than half and that's quite a concern to me. it might
49 make my standard deviations look better but I also think that for the
50 panellist that it's much harder for them to get the sensitivity that they feel
51 they've got on the scale. Whether that's a real effect or not we've yet to
52 determine and that's something we're going to try and do early next year

53 but they tell me that they can't, that they haven't got as much room to
54 manoeuvre. I mean the computer will quite happily convert it into a score
55 out of 150 but it's measuring it in pixels and so there's a reduced
56 sensitivity. The other problem is in the way it works. When the panellist
57 normally taste they have their sheet with all the attributes and they can see
58 where they marked every sample and they can't do that on compusense.
59 They, say if they want to do cocoa in sample 1 and their line comes up,
60 and they mark it and then it goes away again. So if they want to mark it
61 in the same place they can't know that they're putting them in the same
62 place and that, psychologically, is causing them quite a few problems so
63 we're not really sure what we're going to do about that. We're going to
64 write to compusense and say these are the problems we're having and how
65 about future versions being a bit better but I think some of it is within
66 what the computer can physically display and do at one time so I'm not
67 necessarily holding my breath for much improvement on that so for the
68 category scales, yeah it's great. Within 2 or 3 minutes after the panels
69 finished you can get mean and standard deviation, range and you can get
70 them to wait in here and show them... like for a learning thing we've been
71 giving them their sheets as well as putting into the computer so they have
72 a hard copy and that's better for some of their confidence and so they can
73 compare directly what they've put to the mean and compusense will also
74 print that out, have each person's score and the mean in brackets, so that
75 kind of thing is great but for, unfortunately for the most time consuming
76 scale, which is the line scale, it's not that good so far

77 KE

78 Right. You need a higher resolution screen by the sound of things. When
79 we are looking at those final stages of trying to assess whether a panel is
80 ready for project work or not, what's the one thing that you would change
81 which you think would improve the accuracy in the final stages?

82 Expert 2

83 I'm not sure what you mean. You mean in the training process?

84 KE

85 At the end of the training process when you're saying OK I've got a gut
86 feel the data seems right, I think these people are about ready to start a
87 project. What improvements... what's the one most important improvement
88 you think could be made there to increase the accuracy of that decision?

89 Expert 2

90 I think if you wanted to make it a very clear cut decision then we need to
91 have set standard deviations and perhaps ranges, some kind of defined tool
92 to measure them by

93 KE

94 Some of those exist do they not?

295 Expert 2
296 They do exist. They're not always formally applied I mean we don't have
297 it formally defined how many days in a row they should be able to do this
298 or like with duplicate samples, how close they should be, like within a day
299 or over a day so yeah it probably needs a little bit more definition if you
300 wanted to be very strict about it. But also it's difficult because there's also
301 a subjective side to it like if it's just 1 person that's messed up, does the
302 whole panel suffer because of that? and if they don't do it again for a week
303 and then it's another person, do you keep holding them back because of
304 that so it's almost like you need a band with a critical level on either side

305 KE
306 Yes, a red band and then a yellow band

307 Expert 2
308 Yeah, something like that. That would be the most important but there's
309 always the political pressure to keep things going

310 KE
311 Mmm. I guess there's also some psychological pressure on panel members
312 to that if the panel is not put forward for project work then they might
313 think they've failed

314 Expert 2
315 Oh for sure. I've had panellists who have said to me, is there any other
316 panel that has taken as long as us to train? and that kind of thing.
317 Especially if you tell them at the start that you expect them to take a
318 certain time frame and then it takes longer

319 KE
320 Mmm, so essentially one of the rules is don't tell the panel how long it
321 takes?

322 Expert 2
323 Well I think so yeah. I think or be very broad about it, you know 3 to 6
324 months sort of thing, very general because people remember that kind of
325 thing and start to think that they're doing badly and that you might get rid
326 of them

327 KE
328 OK Expert 2, well that's all I wanted to wrap up that particular interview
329 unless there's anything else you want to add on?

330 Expert 2
331 Yeah there is one thing I've just thought of is that I personally, when they
332 finish training and begin project work, tend not to tell them because it

33 creates this great excitement and they're all expecting. I don't know what
34 they expect actually. I think they expect all these weird and wonderful
35 different samples that often are very standard and very routine and so
36 there's a psychological thing

37 KE

38 Avoid the manic flavour behaviour?

39 Expert 2

40 Yeah so like with the chocolate panel they've just finished their first
41 project and I told them at the end, what you've actually been doing is this.
42 And I put up a grid of their standard deviations over the last two and a
43 half weeks and showed them that on one particular day they'd crashed out
44 totally so we'd repeated that and this is what happened and they went, oh
45 why didn't you tell us? but in theory they shouldn't necessarily know when
46 they're doing project work or not

47 KE

48 It shouldn't matter?

49 Expert 2

50 Yeah, it shouldn't matter

51 KE

52 It's like the diamond cutter's apprentice story isn't it? If you make a
53 mistake cutting a diamond you can destroy millions of dollars in value and
54 they have dummy jewels that they train their apprentices on. They cut
55 away on them and then come the time that they've nearly finished their
56 apprenticeship, they say, well today son is the day that you should cut your
57 first real diamond and they go oh but... what if I make a mistake and the
58 master normally says well perhaps we'll let you try one more dummy just
59 to make sure you've got it absolutely right. So they put it in there and he
60 cuts it perfectly and they say well that was your first real gem that you just
61 cut

52 Expert 2

53 Yeah it is that kind of story

54 DURATION 47 MINUTES

165 Full transcript Expert 3 - SESSION 1

1 KE

2 Well Expert 3, thanks for agreeing to do this series of interviews. Perhaps
3 we could begin by you just giving me a general, layman's description of
4 how it is you go about selecting and training people to be taste panellists

5 Expert 3

6 Sure. Well we advertise and as a result of that we get 60 to 80 people and
7 you set them tasks. Well, when I'm running the screening I set them tasks
8 that are related to the product that they'll be evaluating. We go through,
9 first of all, the basic tastes and go through some difference testing and then
10 some difference testing with the product itself using project variables to
11 make different samples and that's supposed to get harder as the work
12 progresses and they get more used to it but, yeah, it's quite difficult. The
13 problem I see is that it's me judging as to whether there's any difference
14 and saying whether they should have got it right or not. I think if I was to
15 screen again, I've done three panels so far, but if I was to do the next one,
16 I'd do a bit of personality work as well, just group dynamics and that sort
17 of thing. So that's screening. And then training starts and you just basically
18 do an introduction to get them all familiar with each other, you know
19 where Susan comes from and all that so they can get down to what they're
20 doing then, yeah, just vocabulary development, that's probably the first
21 stage where they generate lots of words to describe lots and lots of
22 different samples and you try to get samples as varied as possible at that
23 stage. so that they can really get any sort of attributes that they can think
24 of. Then later on, if you're just doing flavour or just doing texture you start
25 to concentrate on those. We probably use the same samples to begin with
26 over again. Then introduce a preliminary, basic scale like absent, weak,
27 moderate, high and try and pick out individual attributes and work on
28 those. And at that stage you probably have a list of 5 or 6 maybe 8 that I
29 think are the same attributes that everyone's talking about and get them
30 all to scale or say whether that attribute is present in a particular sample.
31 So if they all agree that their milky is say similar to someone else's creamy
32 or they all agree that it is present in that attribute then I'll put to the panel
33 that I think that everyone is calling the same attribute different words so
34 what can we call it that everyone will recognise. And you go through that
35 and that's a really circular thing, it goes round and round and they lose it
36 and you have to go back to the beginning again and you know, that does
37 take ages and even now I still wonder... and then you start to work on
38 scaling each attribute with that basic, preliminary scale and then trying to
39 get it to be a bit larger but at that stage a few decisions have to be made
40 by me regard what type of scale to be used and that's always a dicey one.
41 And then it's just basically I find this stage is the longest, trying to get them
42 up to where they understand each attribute and to be able to scale it. And
43 then setting up standards along the scale I suppose is the next stage. So the
44 aim of it is to have them all understand each attribute, understand the
45 scale and be able to recognise specific samples along that scale for that

46 particular attribute. Like it could be like butterscotch in shortbread.
47 Butterscotch is most intense and then weakens, so a standard is needed for
48 butterscotch. And then after we've done that, put them into the panel and
49 put them into the booths and do a few tests to see how consistent they are
50 and then basically that's it. That's really essentially all I do.

51 KE

52 Good. What I'd just like to do is go through that and just ask some silly
53 questions. Obviously coming from a non-sensory background I need to
54 clarify some of the things that I guess you understand because you've been
55 doing it every day. Let me just dig through here a moment and look at
56 some questions. I guess the first one related to selecting. With the people
57 off the street, how do you select the people that are going to go on to be
58 trained?

59 Expert 3

60 The people that come in to be screened, those people are just on our
61 database. They've just shown interest from filling out a questionnaire that
62 we have, they go on to our database and then, yeah, if they're available at
63 the time of our screening and they're going to be available for the next
64 however long then they're screened then based on their performance in the
65 screening test they're put on to the panel. That's if they're 75% correct.

66 KE

67 75% correct on what?

68 Expert 3

69 On the tests that they've performed for the week's screening. In the two
70 panels that I've done they've had two weeks of screening

71 KE

72 Right, and what tests do you run?

73 Expert 3

74 It can vary from ranking tests which involve like where they get a sequence
75 of increasing intensity of sweetness and they have to put them in order,
76 like they might get 8 samples and they have to taste all 8 and put them
77 into order of increasing sweetness. That's one sort of test they can do or
78 threshold test where they've got a sequence of 8 and they have to say when
79 they can taste a difference from the reference... just a detection threshold
80 really, that there is a difference and then a recognition when they recognise
81 what the actual taste is, that's one other test. Difference testing is used
82 which is, well there's several different ones but the common one that we've
83 used is triangle tests which is with 3 samples and you just circle the sample
84 that's different to the other two. The description tests we do with a sample
85 and they have to describe the aroma. That's basically how we do it.

36 KE

87 How come your selection took 2 weeks? You mentioned selection week
88 and then you said yours took 2 weeks

89 Expert 3

90 Oh well we just screened for 2 panels. We only screened for one week and
91 I wasn't involved in doing that screening but previously we've done
92 screening for 2 weeks, it was just a period of 2 weeks and we just did 2
93 weeks worth, but the 2 latest panels, they planned screening for just 1
94 week

95 KE

96 Right, and what are the merits of the 2 options?

97 Expert 3

98 I'm not sure. I don't know what the literature says. I think they do 1 week
99 just for convenience, I think that's basically why. I think in 2 weeks you
00 can get a better understanding of how they taste, perception you know but
01 again when I did it the last 2 times I just did triangle tests and their test
02 results and I perceive that now to have a problem in that now, you don't
03 get any idea how they're going to interact in the group so I'd do some
04 more work on that

05 KE

06 Mmm. So you haven't done any personality screening up till now?

07 Expert 3

08 The last 2 times they did interview them. I don't know what they asked
09 them or anything like that. I haven't interviewed any panellists

10 KE

11 What do you think you'd do if you screening again?

12 Expert 3

13 I think I'd run a training session like with bizarre samples, different sample
14 you know and just put the group together just to see who really takes over
15 or who didn't and all that sort of thing. I'd just do that sort of thing. I just
16 think it's important because I've just learned with the panels I've taken that
17 leaders are definitely the people who develop an influence over the group's
18 performance

19 KE

20 Yes. In what way? What effect do they have on the group?

21 Expert 3

22 Well, I mean they walk in, there's a couple who walk in and go 'yuck', you
23 know, or 'yum' or 'what are you going to give us today'. It just biases and
24 it changes the atmosphere so yeah, I think it's quite important to have
25 people that are really neutral, who are willing to learn

26 KE
27 Or at least have good social skills

28 Expert 3
29 Yeah. That would be the only thing that I would add really. I didn't ask
30 Expert 1 why they chose 1 week but I quite like to go for 2 weeks. At the
31 end of that 2 weeks they've learned a lot but I don't know if that's realistic

32 KE
33 Mmm. So how long would you have them and for how long in the selection
34 process?

35 Expert 3
36 Every day for 5 days

37 KE
38 All day?

39 Expert 3
40 For about an hour. Yeah, but because there's like 60 to 80 people it would
41 be a whole morning. You'd have groups of people coming in and going all
42 the time

43 KE
44 So once you've got your selection process finished, how many do you end
45 up with?

46 Expert 3
47 I'd say 15 to 20 is what I'd choose cos there's always quite a rapid drop out
48 rate in that first 3 months, people that don't like it or don't want to make
49 the commitment, that sort of thing so yeah, and then I'd expect that to
50 drop down to about 12 to 14

51 KE
52 What's the biggest size panel that you'd use?

53 Expert 3
54 The biggest one that we've got at the moment is the bitterness one because
55 they're expecting heaps of people to drop out cos it's just an unmotivating
56 product. When I first took over baked products there was 20 I think and
57 that dropped down to 13 by a year later

58 KE
59 What's the smallest panel that you'd work with?

50 Expert 3
51 To run a panel you have to have 8 panellists. I think Expert 3 works with
52 9 on the chocolate panel and I think the next smallest would be 11 which

63 is the AMF panel

64 KE
65 And what's the basis for using 8 as a minimum?

66 Expert 3
67 Just statistically. You can't work it out parametrically if you've got less than
68 8. Yeah, you need a decent number of people to get it right, that's not
69 really a textbook explanation

70 KE
71 That's OK, I'm not really a statistician. Good, so when you've got your
72 panel of between 8 and 20 say, that seems to be about the range, what do
73 you do with them then?

74 Expert 3
75 Well they'd come in for the preliminary first week and basically the first
76 day is... well probably the first week, we really introduce them to what they
77 have to do and we really push the things like don't wear perfume, don't
78 clean your teeth before you come, you know, you're just reinforcing the
79 principles like just do your own thoughts and all those things. And also
80 start looking at any varied samples. And on the first day they introduce
81 themselves to each other and give a background and I just then start them
82 off writing down as many words as come to mind about each sample that's
83 in front of them. I find they generate so many words, especially if you've
84 got 20, it takes a really long time and yeah, so they'll write them up and
85 then I'll tell them what each sample was and talk about it a bit and then
86 probably the next week, start looking at the words that they've written up
87 there and ask them if they can see any words which could be similar. At
88 that stage you just start varying all the samples that you can think of and
89 if there's one word which keeps coming out a lot, you stick that down on
90 the form, so they've got that to use. So like shortbread we started off with
91 butteriness. Cooked butter flavour kept on coming up so that was worked
92 out and sweetness went down and then baked. A lot of words were
93 generated like baked, toffee, caramel and those sort of things, so that went
94 down. It's at that stage that I get caught up in trying to produce samples
95 that are going to be accurate enough

96 KE
97 Yes

98 Expert 3
99 Accurate enough to get everyone to agree to a particular attribute, like in
100 a baked thing, I mean a baked product, it's really hard to alter just one
101 thing without altering heaps of other things and I find that really frustrating

102 KE
103 So it's a more complex product... or panel training problem than say butter

004 or something like that

005 Expert 3
006 Yeah, well I think so but, yeah, I think the more complex the product the
007 more horrendously difficult it is to get accurate attributes

008 KE
009 Could we just go back to what you were saying about basic behaviours like
010 not brushing your teeth and don't wear perfume, do you have a list of basic
011 skills that you give them

012 Expert 3
013 Yeah, we have got just a general thing on sensory evaluation and it has got
014 things like the tongue, you know, where the stronger areas on your tongue
015 are, not to brush your teeth and that sort of thing. I can't remember
016 everything that's on that handout. It's just really general stuff but on the
017 back of it is a list of things...

018 KE
019 The do's and don't's ?

020 Expert 3
021 Yeah, the do's and don't's which I try and stress to them a bit

022 KE
023 Do you think that I could get a copy of that next time I come in?

024 Expert 3
025 Oh sure

026 KE
027 That would be good. I could just use that for my own reference. It's useful
028 to be able to see all the documentation. By the end of this project I hope
029 to have seen all of the documentation which relates to the training of
030 panels but of course it comes out as it's relevant to what we're talking
031 about. OK so we've got agreeing on words, word generation, looking for
032 similarities and differences and we get an agreed vocabulary. Is that it?

033 Expert 3
034 Mmm

035 KE
036 And then what?

037 Expert 3
038 Then I'll introduce a scale like probably the first scale that I would use
039 would be absent, weak, moderate and strong so they've only got the A, W,
040 M or S for each attribute. You've seen our preliminary evaluation form?

241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283

KE

Mmm

Expert 3

Well it has got the list of the attributes, it has got the scale and they just have to put the capital letters that they think are in there and I write them up on the white board and you'll get variations from absent to strong and someone will say well they didn't think that attribute was in there. You know... why?, and they'll argue that they couldn't taste it. Then I'll ask someone well what was it about the sample that made you put strong and they'll describe that and that other person will still probably say that they still can't taste it but because they're here, I try and get people to interact as to how they're tasting things...

KE

So we've got a verbal description of intensities, that's basically what this step is?

Expert 3

Yeah, and if there are people who have got 'absent' I just say well why did you get that or why did you not get that, or do you think that there's a reason that you didn't get it or that sort of thing, and that carries on. It doesn't actually last that long with that sort of scale. We just try and get them to go through with each of the attributes, then I normally introduce them to a category scale in any of the work that I've done. The primary reason for that is that historically that's what Establishment 1 has used and sometimes I question my logic there, maybe it should be looked into a bit more but... are you familiar with the category scale?

KE

No, I'm not

Expert 3

OK, well it's a 0 to 9 point scale and they can only use the points, it's not a line scale, and the distance between points is not set so the points are anchored, like 1 is the threshold level, 3 is weak, 5 is moderate, 7 is strong and 9 is intense but there's no actual linear, no mathematical distance between the points. So you can say it's between weak and strong but if one averages 2 and the other averages 4, you can't say that it's twice as strong so that's the type of scale that I've used for my work and then basically what I do is I go through different samples, getting them to practice on the scale and just discussion until they've got a good grip of what they're doing and that can take a very long time cos it's quite difficult, it's quite difficult to get the samples that are accurate enough, I mean they may have a bad day in the factory so yeah that just takes a while and that's just basically it. I mean I'm analyzing the results throughout the time and giving them feedback about how they're going. So basically that's it until they're competent at each attribute and we've got a system where, I don't know

84 if Expert 1 showed it to you, where we've calculated the standard deviation
85 about the mean for the panel. If they're too high, if there's too much
86 deviation, they're not ready I suppose, and for their training, they might go
87 and do a panel, like they might have a bad day and their standard
88 deviations are really high, we redo that work where the standard deviation
89 is too high

90 KE
91 Right, and what is the formula? or is there a set of data that you're
92 comparing against?

93 Expert 3
94 No, no, just, well we did a study about 18 months ago and we calculated
95 what the standard deviations had been historically over each attribute for
96 the panels that we had, like milk powder, butter, that we had at
97 Establishment 1. Then we calculated the standard deviation of the standard
98 deviation and worked out realistic, well it's sort of like the quality
99 assurance stuff, you know if it's getting above that then...

00 KE
01 Sort of like an upper threshold and a lower threshold

02 Expert 3
03 Yeah, that's the word I'm looking for, so that's what we looked at
04 historically. With my panels, because there has been no historical work,
05 I've just calculated them myself

06 KE
07 Right

08 Expert 3
09 Just whatever seems realistic

10 KE
11 OK, and what do you judge unrealistic to be measured against?

12 Expert 3
13 Well because it's just a 9 point scale, if the standard deviation is 1, then
14 that's OK by me. On some attributes which are really hard, I take it to 1.5,
15 but it's just that the variation is so high, I'm left not sure if they know what
16 they're talking about... the panel

17 KE
18 So the choice of a standard deviation of 1, say, is your gut feel of what is
19 an acceptable performance

20 Expert 3

321 Mmm. That means that they vary about the mean by 1 so I think you
322 know, that's OK

323 KE
324 How long then does it take to get this panel, you know, we've got to the
325 point where you're doing with categories. Let's just skip my previous
326 question. Where do you go from there?

327 Expert 3
328 Well once they're proficient at using the category scale then they can go
329 and do some project work

330 KE
331 Right, and so your decision to declare them a trained panel is based on
332 whether you believe their performances are acceptable

333 Expert 3
334 Yeah

335 KE
336 And that's what you were talking about with the standard deviations. I take
337 it baked products is quite a new thing. Is that all you look after?

338 Expert 3
339 I do the milk fat panel as well which... well milk fat products which
340 includes butter fats and vegetable oil mixes and margarine mixes for export
341 and just butter

342 KE
343 Right. So none of the group leaders actually do the same panels in effect
344 although people might have done a panel a few years ago and subsequently
345 moved on

346 Expert 3
347 Yeah

348 KE
349 How long then does it take you to get a trained panel?

350 Expert 3
351 Me? 6 to 8 months I think. Milk fat took 6 to 8 months. I mean the
352 literature says 2 to 3 months but to me, I just can't, I just don't know how
353 they get it, I just take 6 to 8 months. The previous panel that I trained, the
354 Madeira cake panel, that took 6 months at least and the panel that I'm
355 training at the moment will probably take 6 months as well. If they've been
356 doing training for ages... I had a rebellion actually! with one of my panels
357 they were just doing training over and over again and they were just getting
358 really sick of it. They couldn't see their way out of this training stuff and

159 in actual fact, I was just waiting for work from these research scientists. I
160 was just thinking any day now they'll give me work and I'll just let them go
161 for it and that didn't happen and so in the end they said, oh I can't
162 remember what happened but they said that they thought they needed a
163 new panel leader because it wasn't going anywhere so that's important.
164 Feedback of their results but also getting real work. They love seeing their
165 own progress. They love seeing graphs of themselves and what the average
166 was and what the standard deviation was and who was too high and who
167 was too low. I mean, they really love that sort of thing. Understanding
168 projects... like I like to explain to them what the project is that they're
169 about to go into and all that sort of thing. We've got a system now where
170 we give them, if they're consistent, regularly turning up, we give them
171 money and things like that which they get quite chuffed about

172 KE

173 Yeah. What causes performance to drop away other than lack of
174 motivation?

175 Expert 3

176 Lack of variation in the samples. Again, I mean if I'm giving them samples
177 which I think... I mean from the processing perspective are different, but
178 are not different enough for them to notice then they find that really hard
179 and they really struggle and they're expecting that they should be different
180 and they can't tell that they're different and they change their profiles. One
181 thing that really frustrates me is the expectation error, in that this is a
182 different number so this must be a different sample and they change it.
183 And you ask them later on 'why did you change your profile?' you know
184 you put in vegetable oil and it was made with butter you know... well you
185 know, it just had to be different. It wasn't but it was and they haven't
186 noticed. There's a lot of psychological error that comes into it I think from
187 just expectations. I've got this theory that there's this really big warm-up
188 effect in both of my products so I try to counter that out

189 KE

190 What do you mean?

191 Expert 3

192 Well the first sample that they have on a particular day is often very
193 wrong, very different and I go back to it and they say oh I'll have to change
194 my profile now having tasted all these and there is such a thing a swarm
195 up effect in sensory evaluation and I think it's present in the work that I
196 do cos shortbread is quite a complex product.....

197 35 mins to here

398 Expert 3 - SESSION 2

399 KE

400 So in the first round of tasks, what I was focusing on was a general sort of
401 description of panel training from the start to when they are declared
402 competent to work on a project. Because that's so broad, what I'm now
403 focusing on... the major area of interest is the last few stages, where people
404 make the decisions, the group leaders are making the decisions as to
405 whether the panels are trained or not. I'd like to focus on that in this
406 session and we'll probably spend quite a bit of time, not only in this
407 session, getting down to some pretty nitty gritty detail about that aspect of
408 things. Perhaps we could just start by you giving me a general description
409 of the activities that you do just immediately up to and how you determine
410 whether a panel is trained or not.

411 Expert 3

412 Well, I... do you want me to go through training, like the samples that I
413 choose, all that sort of thing...

414 KE

415 Just in general terms...

416 Expert 3

417 Well I start off, I just have a vague idea of what sort of product and the
418 particular attributes so I start off and I just get really different, like if I can
419 get samples from overseas that are really different. Like with the AMF
420 panel I start off with AMF and then maybe * so just generate words from
421 that and the aim is to refine that down into specific attributes, is that sort
422 of what you're talking about?

423 KE

424 Yes. So are you talking more about the actual training cycle at this stage?

425 Expert 3

426 Yeah.

427 KE

428 OK. When you start to get a feeling that these people are nearly trained,
429 what do you start doing with them then to actually confirm...?

430 Expert 3

431 Oh OK. Well then I'll get them to profile things and just profile standard
432 samples that they'll probably be getting in project work and then I'll
433 probably, in their results I'll just have a glance at their results and see that
434 there is no-one right out of the ballpark but probably by that stage that's
435 not likely because they've been weeded out to a certain extent. Basically,
436 it's hard to see anyone who's not playing ball in training time but from the
437 profiles I'll do an ANOVA between panellists, and then I'll look at the

138 mean and the standard deviation and if the standard deviation is quite
139 high, then they still need training because there's too much noise around
140 the mean. If there's one person who's outlying on a particular attribute
141 then I'll get them, you know, work with them on that attribute and do a
142 training that's specifically geared up for that attribute so I can see how
143 they perform. I give them some stats analysis like principle component
144 analysis but I've found that lots of those multi-variate analysis methods are
145 too generalised and it's just looking at their consistency really. If I do the
146 same sample twice in a row I see whether the same panellists have given
147 them the same scores.

148 KE

149 Would you do that in a single session?

150 Expert 3

151 No. Over 2 days but that's probably what I'd do and if that all checked out.
152 There's no formal sort of stages that they go through but if I get a feel that
153 I've got confidence in them, basically they'll do.

154 KE

155 Mmm. So how many tests would you put them through, say of each
156 attribute in your product, before you feel comfortable that they're trained.

157 Expert 3

158 We'd probably do each attribute 2 times but yeah, it's quite hard because
159 in the products... it's not so hard with AMF like I'll give them oxidised
160 samples, you know 9 month old samples and I'll expect them to have
161 oxidised and rancid scores and if they've all got that from previous
162 samples, I'll put them in the booths and if they don't find that in their
163 samples then I'll think, that don't know what they're doing or don't
164 understand it but with products like the baked products, to me it's more,
165 it's not even so much in the panel, it's more consistency over time. Like
166 everyone understanding the attributes, understanding the scale... that will
167 give me... it's more of a feeling, baked products, than from the results so
168 it's not each attribute then, so for a sample it would probably be about 2
169 weeks worth of panel, until they're being really consistent.

170 KE

171 So that's, what, 6 sessions?

172 Expert 3

173 Yeah, 6 panels.

174 KE

175 And over that 6 sessions then you said you said you were giving similar or
176 matching samples. So you're going to be able to cover how many different
177 attributes in a single session?

78 Expert 3

79 Umm, I'm not sure what...

80 KE

81 Because I'm thinking, if you're having 6 sessions, and you want to duplicate
82 a particular sample twice, if you're only doing that one thing in each
83 session, that would only allow you to do 3 attributes paired over 6 sessions.
84 I'm assuming that there's more attributes presented in each session.

85 Expert 3

86 Yes there's more attributes, like the thing when I do baked products... like
87 with AMF for 6 weeks you could just change, like do 2 sessions where you
88 just added some oxidised and caramelised and then they'd pick that up but
89 with the baked products, it's more a whole sample because you can't have,
90 unless it's the same batch, you can't really duplicate and between, so you
91 could have a caramelised, you could have shortbread baked with say
92 vegetable oil versus a shortbread baked with butter and the 6 attributes
93 will be different between the two so that's where you do comparisons that
94 they've all consistently increased or decreased between the 6 attributes
95 whereas with AMF I just look for everyone picking up that it's oxidised or
96 caramelised and then consistency in that but with baked products it's more
97 everyone changing the profiles in the same way and also to the same
98 degree on the scale. Cos I don't actually know with each batch of
99 shortbread what attributes are going to change, there's no specific formula
100 with those because of the baking process. You can't just change the fat and
101 say it's going to be vegetable oily now because we've used a margarine, it
102 just doesn't work like that although I'd like it to.

103 KE

104 So what's the critical component that you're homing in on for baked
105 products?

106 Expert 3

107 For baked products, low standard deviation and consistency in the
108 attributes.

109 KE

110 Mmm. How many attributes do a panel normally generate for a baked
111 product?

112 Expert 3

113 Umm, I think about 7, 7 or 8.

114 KE

115 So the figure that they're going to come up with then that we could
116 actually look at the standard deviation of, is that one figure that is right
117 across all the 7, 8 or 9 attributes?

18 Expert 3
19 No no. The mean is for each single one and the standard deviation about
20 that. So I mean in the panel training today, I was looking for everyone to
21 be between 4 and 6 but like there might be one 4, six 6s and two 5s, one
22 7 and that's quite consistent to me but there's not much... yeah that's what
23 I'm looking for but the thing is with the baked products, it just depends on
24 the baking... have you ever seen a binary plot of sensory data?

25 KE
26 No

27 Expert 3
28 It's a sort of like spikes on a wheel and you go so far out, the mean is so
29 far along the line of one attribute, so you have all these different shapes
30 and basically what I'm looking for is that everyone's sort of shape of the
31 profile is the same.

32 KE
33 OK, so you could almost put transparent overlays on them...

34 Expert 3
35 Yeah, of each person for one sample to see whether they agreed for a
36 particular attribute.

37 KE
38 Right, and what's the standard. Do you have a standard?

39 Expert 3
40 The standard is 100% of standard butter shortbread I suppose in the
41 baked products but even they vary from batch to batch just from cooking
42 time, temperature of the oven at the time and that sort of thing so I'm
43 finding it quite hard. I suppose the standard would be 100% butter. It is
44 particularly hard with it, like with AMF it's a bit easier because you do
45 expect the factory to produce standardised... see it's a primary sort of
46 processing. As soon as it starts to get baked...

47 KE
48 It gets hard to control. So all those samples come to you, you don't bake
49 them here?

50 Expert 3
51 Well we bake them up at Establishment 1 but I don't bake them... yeah
52 we've been having just heaps of problems getting consistency between
53 batches.

54 KE
55 Right. So what is an acceptable standard deviation for the baked panel?

56 Expert 3
57 I haven't worked one out yet. Probably it would be under 1.

58 KE
59 Mmm, on a scale of...?

60 Expert 3
61 10. A 10 point category scale.

62 KE
63 So are you using a linear scale?

64 Expert 3
65 No I'm only using a category scale.

66 KE
67 Right and so is there a word related to a number value?

68 Expert 3
69 Mmm. It goes 0 is absent, 1 is threshold, 3 is weak, 5 is moderate, 7 is
70 strong, then 9 and 10.

71 KE
72 Why do you use that particular scale?

73 Expert 3
74 I find it easier. I haven't really used line scales very much. I used...
75 Remember I spoke last week about tradition and it has just been used. I
76 didn't really make a conscious decision I don't think, I just introduced them
77 to a category scale. There are limitations but just talking to Expert 3 and
78 Chris at the time, I thought that a category scale. I think that panellists
79 change line scales into category scales anyway so I just...

30 KE
31 Mmm, they divide up the spaces?

32 Expert 3
33 Yeah, it's supposed to be a continuum but they break it up and think of
34 it that way.

35 KE
36 Mmm. What about line scales. The way you train your panel do they have
37 to select an integer value or can they go between integers?

38 Expert 3
39 They can choose... like 3 is anchored and 5 is anchored. they can choose
40 4 but they can't choose 4.5

91 KE
92 Right, so that's the fundamental difference between that and say line scale
93 where they can just draw a line anywhere. I guess that must make the data
94 a lot simpler to gather.

95 Expert 3
96 Yeah it is simpler. With trained panels I don't think there's much of a
97 difference.

98 KE
99 As I recall, it's because it's the group leader that chooses the measuring
00 instrument. From what I've picked up so far it seems that one of the
01 standards that's being developed within the unit is based on the average
02 standard deviation like that student holiday worker developed?

03 Expert 3
04 Yeah

05 KE
06 So what are the problems inherent in having a group leader who chooses
07 a nine point scale and then another group leader a year down the track
08 uses a ten point scale with the same product and then the data gets put
09 together for an average?

10 Expert 3
11 I don't know but the way I envisage it is if I were to change from a 9 point
12 to a 10 point scale and the acceptable standard deviation was 1 in 9, then
13 I'd just take that percentage and apply it to the 10 point scale so that
14 would work out to be 1.2 or something. I think I'd do it like that. I think
15 I'd have to talk to a statistician to work out whether that's feasible but
16 that's the way I'd transcribe. To me I couldn't really train these guys on a
17 line scale without going back to basics cos it's a different fundamental
18 concept.

19 KE
20 So you expect to keep this panel on the same measuring instrument the
21 whole way through?

22 Expert 3
23 Yeah. In some ways, I mean I tossed up whether to use a 15 point scale
24 but 9 point seems to work quite well. 15 point scales are being used a lot
25 in sensory but I don't know whether...

26 KE
27 Just getting back to the stage where you are testing for competency and
28 putting them through duplicate samples in each attribute, at what point
29 would you say, these people haven't got it?

30 Expert 3
31 This person or these people?

32 KE
33 As a panel cos you're trying to produce a panel for project work aren't you
34 so...

35 Expert 3
36 If I had the training and they're all over the place and they don't
37 understand the samples then I'd say come back tomorrow and I'll see if I
38 can get more... like we have the training and they might not go very well
39 on one particular sample and it's quite an important sample or something
40 and I'll think I'll get them back next week and we'll still have this sample
41 perhaps with a different angle, so that's what I'd do first and if they still
42 can't get it I'd just take them back to basics again and start them on a
43 really wide variety of samples and then try and tune it down again and if
44 that doesn't work I'd probably go to Expert 1 and cry my eyes out. So
45 that's sort of what I'd do.

46 KE
47 OK. I'm just trying to get a feel for the time scale so you said 2 weeks so
48 6 sessions would be the minimum, is that normal?

49 Expert 3
50 No. Because the baked products panel hasn't really been in panel work yet
51 and it's sort of a wee way away going to the AMF panel, you know I did
52 a couple of panels and if they went alright on those then I'd think OK,
53 then do another sort of block of training and then do a couple of panels
54 and see if they've got those attributes sorted and so that would probably
55 be a week, 2 no 3 weeks in blocks I did with them. I didn't actually
56 specifically at the end of their training do 2 weeks of competency testing
57 to see whether they were consistent.

58 KE
59 Right, so it's just woven into the normal training cycle as you go along?

50 Expert 3
51 Yeah. And then 6 weeks down the track after you've done 2 more blocks
52 you might go back to the block that you did 3 times ago and see how
53 they're still going with that, that's more realistic.

54 KE
55 Right, and for a panel, in the worst case scenario, who just aren't getting
56 it, and are you saying we'll abandon the panel then? I'll try and say that a
57 bit better. We'll abandon the testing stuff in the booths and go right back
58 to training, does that decision ever get made?

59 Expert 3

70 Not really

71 KE
72 So you keep on this rolling week of a couple of sessions in the booths and
73 then a week of training?

74 Expert 3
75 I don't tell the panel what to taste sort of thing. If they're saying to me, we
76 just don't understand it then I sort of drop it more than anything else.

77 KE
78 You drop the attribute?

79 Expert 3
80 Well drop the whole tack that I'm on at that time, try and change tack and
81 maybe try and go around the particular attribute that I'm working on. To
82 me it's just that they're not understanding it so I'll try and change the way
83 I'm teaching it rather than think, it's not worth it. Because when they set
84 up their words at the beginning they seldom change so I try to, if a sample
85 has come in for a project and they're really strange then I'll give them to
86 them... I suppose I, yeah basically I just change my tack, does that answer
87 your question?

88 KE
89 Yeah, I can understand that. If you get one attribute which is continually
90 throwing them, say the baked products, they can reach agreement on 6 or
91 7 others and 1 consistently causes a problem, what do you do with that?

92 Expert 3
93 That's hard because I suppose that's like feedy in AMF. There's this typical
94 NZ flavour in NZ dairy products and because it's a NZ panel they can't
95 pick it up because they're not familiar with what overseas people taste and
96 if I can't get overseas samples for them to taste then they can't really pick
97 up that attribute but then scientists at Establishment 1 will say but what
98 about this NZ flavour, why haven't we picked that up? So they've said to
99 me, this stuff DMS, Di methyl Sulphide is a component of the feedy
00 flavour so I added that to the samples and they said it smells like the back
01 end of a cow basically, and I sort of talked them through it and talked
02 about NZ flavours and we called that feedy, you know it's the smell of cow
03 but they'd only scored that once in any experimental samples that we had
04 because it's been added. Apart from that, if an attribute comes up, there's
05 not many attributes that do, well I haven't got any attributes that do... like
06 with the madeira cake panel there was an attribute called whey, which was
07 very hard for them and it just took ages and ages and I just basically kept
08 training them and training them and changing tack and repeating things
09 trying to get things that can taste and stuff, to understand it. I reckon the
10 essence of a good panel is each understanding what the actual flavour is.
11 A lot of people just sort of agree and they don't actually pull out that

12 particular dimension of the profile, I can see it in their faces when they're
13 sitting there, generating the words. They think this is a shortbread, what's
14 shortbread got in it, oh it's floury, you know, if you can't actually taste
15 floury, you know what I mean? And it's like when they... I'm rambling, do
16 you want to...

17 KE

18 It's OK...

19 Expert 3

20 There's fat that's added to a vegetable base and they say, himix, it tastes
21 like himix. They've never tasted himix before but they just know that that's
22 what's added, you know, so they don't actually describe the flavour. I sit
23 and I try to tell them that that's the wrong thing to do, well it is wrong
24 because they're not describing the flavour. It's not going from their tongue
25 to their brain, it's going from their brain to their tongue and I find that
26 really hard. Their minds really do jump the gun a lot, so I'm continually
27 trying to get them to understand the attributes, and I think that's the
28 essence of it all.

29 KE

30 So what happens when a project is stacked up and waiting and you can't
31 achieve agreement amongst the panel, what do you do then?

32 Expert 3

33 Sometimes I'd put them in a panel just to see how they'd go and I've
34 probably trained them with the samples that they'll be evaluating anyway
35 and if they just way out of the ballpark on them then I'd put it off
36 probably. But yeah, that has it consequences but rather than get results
37 that are quite fuzzy, I'd rather just wait. That's what we did with AMF
38 actually, I thought they'd be trained in October and they weren't trained
39 until January. The project started and they just had to wait with the
40 samples in frozen storage until they were ready.

41 KE

42 Have you ever got to the stage where you would declare a panel
43 untrainable?

44 Expert 3

45 No, I haven't but sometimes I wish I could. I don't think that there's a
46 product that's untrainable but I think, yes, there are panels that are
47 untrainable. No, I think, I just can't get these people to come together, the
48 reason is, it takes a long time to come to that understanding and in that
49 time there's been a lot of money poured into them, so to me I feel sort of
50 committed to think oh well, I've got to work with what I've got, rather than
51 dump the lot and try again. To me also, you can get hung up with that
52 because in fundamental science if you find it doesn't work out then you just
53 change the method and realistically that's probably what we could do.

'54 KE
'55 Right. And what would you change?

'56 Expert 3
'57 Well, I learned a lot when I started training down the dairy factory and I
'58 learned heaps about organising, what's important in a baked product and
'59 did a lot of... you know, cleared my own cobwebs out. I didn't have any
'60 idea really how the panellists would evaluate the samples, how the samples
'61 would cook, what I needed to do to duplicate them, all that sort of thing.
'62 Now I understood all that, and in fact with shortbread it has been a lot
'63 easier, it's a different product but it has been a lot easier to train... it's sort
'64 of like steps along the way and I think that you should head up mini
'65 projects in training a panel and if you can break it all down into little
'66 decision points that you have to make along the way, which is probably
'67 what you're going to be doing for us, then you are that step ahead. I mean
'68 I had an empty head about madeira cake I just learned it as I went along
'69 rather than actually planning what I might learn from the beginning.

70 KE
71 Right, and a lot of it comes down to personality at the end of the day as
72 to how much grunt you can get out of the group, I suppose, and your skills
73 as a facilitator.

74 Expert 3
75 Yeah, sure.

76 KE
77 Have you ever struck a group you didn't like?

78 Expert 3
79 Yeah, I suppose. One milk fat panel out at Establishment 1 I found really
80 hard. I mean I was horrendously nervous when I first went in there and I
81 don't think I one over their confidence and I just felt like I was on the
82 back foot all the time and I didn't want to be and I didn't really
83 understand what they were doing and that sort of thing and that was
84 basically because I didn't really understand what I was doing. Yeah, so I
85 suppose I have. There's panellists you don't relate to, you don't understand
86 where they're coming from which is really hard and they're quite often and
87 they're sort of outliers as well. There's one person on baked products who
88 is an outlier consistently and I don't know whether I've conditioned
89 everyone else and she won't be conditioned by me or whether she's an
90 outlier.

91 KE
92 Or a combination

93 Expert 3
94 Mmm.

'95 KE
'96 What would happen if you had a panel that consistently wasn't performing.
'97 You said before that you'd go to Expert 1 in the last instance. What would
'98 she do?

'99 Expert 3
'00 If I had a panel that was consistently not performing I'd go to Expert 1
'01 and I'd explain to her. What would Expert 1 do? Expert 1 would probably
'02 ask me if I'd done a lot of things, suggestions to try and get them back into
'03 line and I'd probably go away and... I'd to be able to make the decision
'04 that they're worth it or they're not worth it but I'd have to make that
'05 decision with Expert 1 based on my recommendations I suppose.

'06 KE
'07 Yeah. What does it cost to get the panel to that stage of testing?

'08 Expert 3
'09 Thousands of dollars say for baked products to get them going, say
'10 \$16,000.

'11 KE
'12 That's project ready?

'13 Expert 3
'14 Yeah. They were project ready and then they decided that they weren't
'15 going to do any of that project work. With madeira cake, yeah that would
'16 cost about \$6,000... \$16,000 sorry.

'17 KE
'18 \$16,000, yeah so it's a big decision to say they're not ready

'19 Expert 3
'20 Yeah, it's a big decision where you spend \$8,000 and you just don't think
'21 they going to come together to say, no, we'll start again.

'22 KE
'23 Mmm. So what's the cost of using them on the project if they don't
'24 perform?

'25 Expert 3
'26 Good question. With previous used panels like the butter panels and stuff,
'27 I think you can work that out but with panels that you haven't set up
'28 before, I don't think that you can.

'29 KE
'30 Just getting back to the testing phase if we can call it that when you put
'31 them in the booths, what are they presented with?

332 Expert 3
333 With AMF, I presented them with samples with what I thought were
334 varying degrees of oxidation and caramelisation and rancid, they're all
335 doctored by me and, they're all made up by me I mean and based on
336 previously what they've found to be significantly different. There's a lot of
337 feedback involved, what they say, what we do and so that was what I based
338 those samples on. With the madeira cake, also with AMF I try to get real
339 samples like I try to get things that have been rancid for 8 months, and
340 then something that's moderately oxidised and put them in the booths and
341 see whether they still say that they're moderately oxidised. It's sort of not
342 doctored samples because my interpretation of doctoring might be wrong
343 for that sort of thing so I try to make them as real as possible. With the
344 madeira cake panel I just varied the ingredients and in a lot of the ways
345 , it's probably not what they would have got in a project but that's just so
346 that they get different samples. My biggest worry was that project samples
347 were not going to be significantly different enough for them to pick up the
348 differences, so that was basically what I was doing there.

49 KE
50 So what's the reference point for them?

51 Expert 3
52 The references are in the definitions that we've developed in the training.

53 KE
54 Right. So there's no actual reference sample that we can taste, or you can
55 test that their reference hasn't moved. And that's the same with
56 shortbread, the reference is in the definitions of the reference samples that
57 we've defined and the definitions of the term and you see that's when... I
58 mean 2 months ago, we might have defined the terms and definitions and
59 everyone might have felt really good about them. 2 months down the road
60 there's a new box of butter, it's baked differently and makes samples that
61 are different anyway so I mean that's where I find it really hard.

62 KE
63 Are there other products that can have reference samples?

64 Expert 3
65 With AMF I have a reference base which is a freshly produced AMF from
66 bay milk products and I do things to that like 24 hours under oxidising
67 lights or there's another fresh caramelised sample that I get from
68 northland, so yeah it's more standardised cos factories can produce to
69 specifications but even that can be quite iffy.

70 KE
71 So, on any given day, this baked products panel is going through how many
72 different samples?

873 Expert 3
874 4

875 KE
876 4 samples? And what combination of things are you looking for or what
877 combination of things will you present?

878 Expert 3
879 Well today in the training I had 100% butter sample, a sample that was
880 made of a vegetable oil and milk fat mix, a sample that was made with
881 acid* butter, butter that has just had acid* added to it and a sample that
882 was undercooked. The same formulation as the one with the lactic acid in
883 it but it was less cooked and that's what I would give them in training and
884 in the booths.

885 KE
886 Right, so you choose that combination of products and they bake it for you
887 out at Establishment 1?

888 Expert 3
889 Yeah

890 KE
891 Why those particular combinations?

892 Expert 3
893 Well what I thought when I went in there today was if standard butter was
894 a typical thing that they'll be used to, they've had standard formulations for
895 a long time, and I expected them to recognise that as standard and then
896 the three samples as experimental cos they had different flavours in them.
897 So the next time I made up a spec 219 that I expected them to find less
898 butteriness and maybe vegetable oil flavours and a different cooked flavour
899 and the next one, the acid butter, I didn't quite know what they'd do with
900 the butter flavour but I expected them to find a different cooked flavour
901 again and then I thought they might be able to get the lactic acid flavour
902 in the uncooked and see if they could get it in the cooked cos it's still
903 there, but I thought it was more pronounced than uncooked formulations,
904 so that's basically what I went in there expecting. They didn't find that at
905 all, they said... oh they essentially did but not like I thought they would. But
906 yeah, talking through all the attributes of the baking and stuff, we came to
907 some agreement. See the thing that I find the most frustrating is next week
908 if we get those same samples baked again, I'll give them to them and
909 they'll say, nah, they're not the same as last week and I don't know
910 whether to say to them, it is the same as last week or whether... even
911 holding it over a week, they'll say, no they've changed something that week.
912 And realistically they could have so it is really hard.

913 KE

014 Yes. Are you talking about a training environment round the round table
015 at this stage?

016 Expert 3
017 Yeah

018 KE
019 So once they actually get into the booths, particularly with the computers
020 there, they'll get standard deviation feedback straight away?

021 Expert 3
022 Yeah

023 KE
024 So they're going to get that information eventually, the stuff you're
025 hesitating about. Is it the same, will the computer feed that back as well?

026 Expert 3
027 Same as what?

028 KE
029 Tell them what the samples were?

030 Expert 3
031 No. They'd have to go back into the booths and I'd tell from that. So those
032 are the sort of samples I get, I mean when I write out my baking list I try
033 to get the samples baked at this stage I don't really know how diverse the
034 samples that are going to be baked for the project will work out. But what
035 I try to do is typify to them each attribute that's on those lists and it been
036 really hard to just change one attribute without changing all the rest so
037 now what I've decided is I'm just going to try and do everything and just
038 make sure that they've gone over and over each sort of baking combination
039 but that just means heaps of training.

040 KE
041 Mmm. How many different baking combinations are there say for
042 shortbread?

043 Expert 3
044 Hundreds

045 KE
046 And you just basically choose before...

047 Expert 3
048 Most relevant changes. I mean I could bake them for 2 hours at 50 degrees
049 or something like that but that's a waste of time. But getting a realistic
050 niche of the variation that's going to be in the project work that I'll be

51 doing and then being able to relate that to commercial samples as well.

52 KE

53 OK, so how many are you likely to present them with for testing to see if
54 they're ready for a project. You're going to repeat each one so how many...

55 Expert 3

56 Probably a week, I'd say a week of panels going into the booths, coming
57 out and training straight after that and probably be a week with the
58 samples that they'll probably be using in the project plus some of the old
59 references that they'll be familiar with. Like I'd have, if Russell said to me
60 he was going to do a baking trial with different amounts of this lactic acid
61 flavour I'd start them of the week and the first day I'd probably do two
62 standards, two whole milk formulations, whole butter formulations and two
63 with the lactic acid amount that they're used to from training.

64 KE

65 That's their 4 samples

66 Expert 3

67 That would be their 4 samples. They'd probably be duplicated so that I
68 could see if they were consistent within each person and then I'd take them
69 back into the round table room and make sure that everyone recognised
70 that the lactic acid, well the different flavours that were in the lactic acid
71 were more or less standard and if everyone was happy about that then the
72 next day I'd probably have the standard one again, and perhaps the highest
73 formulation of lactic acid that Russell would want and then the lowest
74 formulation of lactic acid, see how they go on that, then the next day I'd
75 probably put in a standard and then a duplicate of one of the experimental
76 samples, see how they go with that and just another sample, probably from
77 the project. That standard might be the lactic acid one they were used to
78 or the whole butter formulation and I'd just see how they go over that time
79 and if they need it on the second time round, I'd probably have another
80 training on the next day and see how they went with that and if they got
81 that I'd put them back in the booths and perhaps do that same panel again
82 or just vary it a little bit. That's's basically how I'd gear them up for
83 project.

34 KE

35 Right, and you'd keep on repeating that cycle until you got a satisfactory
36 standard deviation?

37 Expert 3

38 Yeah

39 KE

40 So it's actually quite a small number of baking combinations that you're
41 using.

092 Expert 3

093 For project specific work, yeah

094 KE

095 What if you were training a panel where you didn't know what the project
096 would be?

097 Expert 3

098 Well that's what I'm doing now and that's just trying, what I've tried to do
099 is suss out the parameters of shortbread flavour, that's what we've been
100 doing. I've worked with the flavour chemist, Russell, and he's been trying
101 to enhance specific attributes within the shortbread and I think that they're
102 just starting to understand specific attributes. Like when shortbread bakes,
103 there's 3 types of flavour that can develop. The caramel flavour, a
104 butterscotch flavour or a toffee flavour and a lot of that has got to do with
105 the degree of browning in the biscuit but for each person to understand
106 each of those attributes and for caramel, the most intense is caramelised
107 condensed milk then they come back down the scale. Shortbread is
108 probably only ever going to get up to about 5 but that's where I try and get
109 a 5, then I try and get a 3 and then a 1. Absence is really easy to find and
110 strong is really easy to find but it's making sure that they know those
111 grades in between. Then, like with the butterscotch, 9 on our scale is
112 Heard's butterscotch and comes back down and that's your most typical
113 baked flavour in shortbread, so that was really important to get them to
114 evaluate that accurately and the toffee flavours as well. So what I've been
115 trying to do is get Russell to give me samples that are along that line,
116 along the butterscotch line or along the toffee line.

117 KE

118 Right, so you can just concentrate on the one attribute at a time

119 Expert 3

120 Yeah, but in doing that, quite often I'll think, good I've got this
121 butterscotch sample, and then they'll say, yeah, there's butterscotch there
122 but the butter sweetness has changed and, you know what I mean? There's
123 so many things going on that that's really hard to get them to... I'll like
124 them to just switch off and just talk about the butterscotch but they don't.
125 They start saying, yeah it's so much sweeter than that first one sort of
126 thing, so yeah, it's really hard so I get that feedback that the sweetness has
127 changed and I go back to Russell and... There's some commercial samples
128 that I've given them that have got totally different flavours than the
129 samples that Establishment 1 produces so I try with that to work out the
130 other flavours, I've put a category in which defines a lot of other flavours
131 but choosing just that bracket of samples of shortbread... see I could bake
132 one that was different but that's ridiculous, sort of thing. But there is a
133 potential flavour in shortbread that's ... so I suppose there is a lot of
134 shutting down from a range of samples that I didn't test for.

73 Expert 3

74 Yeah

75 KE

76 And I guess at the other end, It gets fairly open ended as to how long it
77 takes to achieve competency on every single attribute, or does it?. Is there
78 a figure that you have in mind at which point you have to change your
79 tack?

80 Expert 3

81 Where they're not competent? It's just from their comments like if I've got
82 a sample and they've put up their profiles and on one particular attribute
83 it goes from 0 to 8, I'll say to the people who put 8, why have you put it
84 so high, please explain to everyone else. They'll explain it the way they
85 thought it, and then I'll say to the person who put 0, now can you relate
86 to anything that person said and they'll say yes or no. I'll say why have you
87 got 0, what was it about the sample that made you think that there was
88 none of that particular attribute. If they can't agree then I'll ask everyone
89 else if they can understand what's going on and if they can't understand it...
90 I mean quite often when it's 0 to 8 it means they just don't understand
91 what the attribute is. Cos I used to have a cooked butter and an uncooked
92 butter flavour in shortbread and uncooked butter was consistently all over
93 the place. They'd say they could taste uncooked butter and when I got
94 them to talk about it they didn't know what they were talking about, they
95 were just talking through a hole in their head so i just took it off. Probably
96 what I would do is just take it off. I said to them, if I can find a sample
97 that typifies this, then we'll put it back on but unit I have a sample that
98 you guys can consistently find uncooked butter flavour, then I'm not
99 putting it on. So that's basically what I'd do in that situation.

100 KE

101 Good. And so given the nature of your pattern, that you roll them in and
102 out of training with the booths I guess the question of how long do you
103 keep going in the testing phase is less relevant because you don't actually
104 have a discreet period where you just sit them in the booths continuously,
105 do you?

106 Expert 3

107 No

108 KE

109 In your experience, how long do they tend to go on for, from the time that
110 you sort of get a gut feel that they're about ready, how long does it go on
111 from that point to you saying they are trained?

112 Expert 3

113 I take ages to train my panels. Probably 6 weeks

14 KE

15 Why do you say you take ages?

16 Expert 3

17 I don't know. It just takes me ages. I don't know whether I'm not pushy
18 enough or I don't understand them enough, but it just takes me ages. I'm
19 really confident once I get them trained that they understand everything
20 but to me, like they'll just say I don't understand, I don't understand and
21 I try to get them to understand why the panels changed or the flavour has
22 but if they can't still understand it at the end of the training session, I leave
23 it and I think, well it's up to me to find samples to make that person
24 understand that attribute. Therefore sometimes it can just go on for ever
25 and one person just doesn't understand one attribute. Eventually they get
26 there and that's really good but, yeah, I'm just not into changing... I should
27 sit in on a few other people's training sessions and see what they do, but
28 I just don't see any way of changing someone. You ask them, why they're
29 0 and there's one person with 5, why did you get 5... because I could taste
30 oxidised, and I mean, they did taste it, you can't just say, sort of thing, and
31 then if you get them to change that one in comparison to that one and
32 they still can taste it there, I don't know, I just can't... I've never said to
33 anyone that they're wrong

34 KE

35 Mmm. I guess alternatively there is groupthink where the people who taste
36 a 5 will come into line with what everyone else says regardless, and that
37 could be wrong

38 Expert 3

39 Yeah, but that comes down to the fact of understanding it. Remember I
40 told them there's none there, because everyone else has said, they don't
41 understand that so they've got to understand each attribute and each scale
42 which I think they need to learn themselves. They have to learn as a group
43 but individually with the samples, do you understand that?

44 KE

45 Well thanks Expert 3. That's all I have for today.

46 **DURATION: 45 MINUTES**

Appendix C

Transcript Excerpts

1 ... We get them straight off the street. We select them on the basis of how good
 2 they are at tasting a particular product and in order to do this we employ
 3 threshold testing which is basically how good they are at the four basic tastes. We
 4 need to find out how good they are at discriminating between certain products for
 5 example a cheese panel gets to try all cheese products. We also use triangle tests
 6 where 2 samples are the same and 1 is different. They need to be able to spot the
 7 difference. At the end of 1 to 2 weeks we eliminate about 75% of people. That
 8 usually leaves about 15 to 20 people. Now some of these people are good at one
 9 product but not at another. Then in the first few training sessions after selection
 10 we get them to describe everything about the product for instance flavour, texture,
 11 colour. We take all the words that are used and the group leader finds matching
 12 words to get a shared vocabulary. That usually takes a couple of days. Then we
 13 introduce intensities. The group leader works out some end-of-scale words like
 14 weak, slight, strong, and then moves from those onto numbers. We would
 15 probably spend 2 to 3 weeks on words. In actual fact, the process of converting
 16 words to numbers is easy by tying values to specific words. Then the pressure goes
 17 on the panel because this is when the leader sorts out the outliers. We can spend
 18 1 to 2 months getting the numbers right. We also start measuring variation,
 19 analyzing the standard deviations and giving that feedback to the panellist. At the
 20 end of this process we make the decision as to whether this panel is fully trained
 21 or not... Once a panel is trained, they are 100 times more sensitive than a person
 22 on the street to that product. We've got to try and get rid of the variation between
 23 the experts at that decision point as it's critical to the quality of the panel

24 KE

25 Have the group leaders been given feedback about their variation in the same way
 26 that the panellists do?

27 ... No... there are definite differences in assertiveness and how hard each one
 28 pushes a panel

29 ... There are standard deviations for each product based on historical data but
 30 leaders aren't using them so my question is what are they basing their decisions
 31 on? You see the length of panel training ranges at present from 4 to 6 months,
 32 depending on how motivated and keen they are. It also depends on how good the
 33 leaders standards are and how pushy they get

34 KE

35 Right. So how many hours training per week do they do?

36 ... 3 times per week. In training that is a 45 minute session ranging up to 1 hour

37 ... During screening they get \$5 per day plus a \$5 bonus to be there at the end.
 38 Normal pay for training is \$10 per session regardless of the length of time that
 39 they're there

40 KE

41 Mmm. What causes the decay in performance?

42 ... Changing lifestyle, pregnancy, stress and tiredness. Fatigue and adaptation due
43 to over testing... When we give results back, we hide the names except for those
44 who are consistent outliers. What I want from this is something neutral and
45 standard for group leaders to refer to

46 ... It all starts I guess with interviewing. The first interview that you're ever likely
47 to do is going to be on the phone and there's normally a standard sheet that you
48 go through, or if they don't want to do it on the phone, you send it out. So the
49 basic information that you've got on a panellist is what's on the sheet like their
50 sex, what they like or dislike in food, whether they're allergic to stuff, how
51 committed they are. When they come in for screening, toward the end of the week
52 we actually interview everyone and we ask them, from another standard sheet,
53 what their commitment is, are they going to get employed, how happy they feel
54 about coming in. What we're trying to ascertain is if they are going to be a good
55 group person, whether they're going to leave us in the lurch 2 months down the
56 track and whether or not they're actually interested in it. We try to find out
57 without being too subtle whether they're doing it for the money only or whether
58 they're actually genuinely interested... I try to do is have 2 or 3 people in our
59 section sit in on the interviews

60 ... sifting through and sorting out the ones to be included. I guess you could call
61 that the personnel selector. It overlaps with the interview but is just a
62 confirmation. Have we got a good group person in there? But the main selection
63 criteria is still the results from their screening tests. It's based on the most
64 objective information that we can get so in combination with their results from the
65 screening tests and the limits set on them, plus the interview, that's how we would
66 select. We tend to select more people than we need because they usually drop
67 out. If we were to select just on the basis of results, we wouldn't get enough. We
68 often drop the standard to increase our numbers

69 ... You've got to take a whole group of strangers, people that don't know each
70 other and we've got to convert them into a group of people who know each other
71 and trust each other enough to say well I really don't think that it is such and such
72 and feel comfortable with that. So this is where your skills are really brought to
73 bear as a facilitator

74 ... The first few sessions are really important ones as they set the scene so we've
75 tried to do some fun things and some interesting things. Especially introducing two
76 groups of people like where we are reinforcing a panel and you have to bring the
77 new lot and the old lot together... we've done a few little things like blindfolded
78 a new one and the old one feeds them a product and they have to say what they
79 think it is. Then they whip off the scarf and introduce each other. You have to be
80 able to read people and know when to push and when not to. One of the harder
81 roles is knowing when to push them. I guess that leads into the next task which
82 is as a trainer.

83 ... you've got to analyze the performance, looking at questions like why is the
84 person doing things wrong, what can I do to improve them, what steps do I need
85 to take to ensure that they do pull up, and if they don't making the decision to get
86 rid of them. It's analysis I guess.

87 ... It also comes in when you want to throw a bit of stats around the place. You
88 want to try and link something that's happening with the panel with something
89 chemical or physiological or whatever but analysis comes further down the track.
90 It's when things get tough that you try and get a lot more discriminating and
91 analyze a lot more

92 ... the panel leader has to make a final decision about the panel or the panellists.
93 We say that the panellists have got this apprentice period. Now I could be really
94 hard nosed and say right, if you're not performing in 3 months, then you're down
95 the road but lots of other issues come in and effect that. Like we have to keep
96 persevering with useless ones to keep the numbers up or they are going to bad
97 mouth us if they leave

98 ... There's also decisions like are they ready now or are they not. Other issues that
99 affect that are if the project work is coming, or getting nagged to make some
00 money out of the panel

01 KE

02 Can a person that is socially disruptive bias the tasting?

03 ... Definitely, both in the results of the panel overall, and also in the motivation
04 of the other panellists

05 KE

06 I see. What I'd like you to consider now are some of the key items that make up
07 the tasks we have discussed. If we can look first at the task of interviewer. What
08 do you feel are the elements that make up this task?

09 ... the most important ones would be... articulation, the ability to speak well to the
10 applicants. Social skills are also important and we also have a checklist that we
11 work to

12 ... I've got we used to three or four months so after they've been going sort of two
13 or three months you used to start looking quite carefully at their individual results,
14 ok. So we would actually start using a calculator or putting it through the
15 computer often we had at the time and getting out the mean and getting out the
16 standard deviations and having look to see if that standard deviation is coming
17 down

18 ... just before we consider them fully trained we do what we call half blind tests,
19 where we take them through the training situation, put them in the booth just to
20 get them used to that then after they have done the testing in the booths you take
21 them into the round table room then and you right their results up on the board

122 ... probably do two or three weeks of that on perhaps a bit of a mixture of project
123 work and a mixture of standards to see how the variations have gone. And then
124 if that hasn't worked out well you give them another shot at the training

125 KE

126 So this decision, is that based on one or more half hour sessions?

127 ... It is definitely based on more than one. See some of them likes of Expert 3
128 she's been running a lot of half blinds. Normally we probably run two or three
129 weeks it really depends on how good they are, if they're not any good well you
130 whip them back in and work them back in the half blinds again but whilst they're
131 not two or three weeks, it's just until they feel comfortable working in the booths

132 ... to get their confidence up and assure them that yes when they look at their
133 results on the board they are still within, still even with the rest of them.

134
135 KE

136 So you are looking at six to nine sessions

137 ... Yes, you're really evaluating how good they are and it's just a chance, we
138 usually have a look at them and then get used to the actual project work, it would
139 be about two or three weeks

140 ... varies depending on the product, a couple of years ago we did a study, because
141 the literature doesn't give you much guidance on how to check panels variance,
142 it was really a measure that we decided on that we could effectively use that was
143 efficient and it was something that we could do easily so we did a study, we got
144 a summer student to collate all the results from every product that we had and
145 work out the average standard deviation because there was no point in going for
146 something that was too low or wasn't even reasonable and because we use
147 different scales for different products we had to work out standard deviations
148 based on a line scale and a category scale wouldn't work, so she found that on
149 average that the really average attributes like sweetness or something that they do
150 every day, it could even go below one okay, we found an average standard
151 deviation of .5 for instance for sweetness in milk powder which, I mean that is
152 almost perfect given a group of people. But some of them where over one on the
153 ten point scale so you know you're looking at ten percent fuzziness in the data
154 which is quite reasonable really, so it depends on the attributes, if they are not
155 very common attributes then they are very difficult doing and standard deviation
156 might be something like 1.5

157 ... eventually we should have standard deviations for every attribute

158 ... some of the new panellist that are trained at the moment, they are coming up
159 with words that we have never used before so we have got no history in which to
160 gauge standard deviation but depending on the scale KE, like I said if it get any
161 where near two on a ten point scale it is a problem no matter how much history
162 you have got to base it on

163 ... if the attribute is always causing a variation I would get rid of it, because it is
164 obvious not enough people are using it and can relate to it in order to get that
165 variation down

166 KE

167 What system do you use in the panels then for the panellists to record and report
168 the scores about the product in use?

169 ... what happens in the training process they start out with a blank sheet, they
170 move on to a sheet that's got a lot of words on it and those words are eventually
171 cut down to what we call our final profile. Then it's on those words that they do
172 their scoring, be it on a line scale, which is chocolate at the moment, be it on a
173 nine point scale which is AMF, be it on a ten point scale which is most of our
174 products, category scale. So your panellists are either using numbers or they are
175 using a line scale with which to do their scoring and it's from that that we then
176 convert that into a mean score with a standard deviation so at the end of the day
177 the panellists don't get instant feedback as to what the standard deviation is on
178 that day but that will change once we've got our computers in, that's the whole
179 idea, instant feedback. If you can say to a panel that day that they're way out of
180 kilter, I mean two weeks later it's no good hearing that news they can't do
181 anything about it then, which is what's been happening

182 KE

183 Oh, I see. Why do the group leaders, why do some use a line scale, some use a
184 nine point and some use a ten point scale?

185 ... some people have an aversion to numbers. Any scale that you use, there will
186 be end-of-scale avoidance, so your idea is to try and get a scale that's wide enough
187 to cover all possible options but also account for end of scale avoidance. So if
188 you've got a five point scale you're only effectively going to get three numbers
189 used because everyone will avoid one and avoid five. So that's one of the aims of
190 our ten point scale, that has enough, gives you enough discrimination and, for
191 them to avoid zero and ten but you're still getting enough discrimination on the
192 rest of your points that are left. Now Expert 3 uses a nine point scale, I'm not
193 quite sure... her and I have had long discussions as to her using a nine point scale,
194 so you should ask her about the theory behind doing that. A line scale is meant
195 to get rid of this apparently because with the category scale like a number scale
196 the assumption is made that the difference between one and two, and the
197 difference between two and three is the same. It's not actually true

198 ... what we've used is what we call a 150cm line, and we have intervals at 10mm
199 and 140 or whatever it is. And then, so you work out in terms of 150cm but, yeah,
200 you convert that basically into a percentage

201 ... when you start interchanging the scales with the product we're going to run into
202 problems with the standard deviation

203 ... On any sheet that you look at, there's room for comment and these qualitative
204 people will always write something...

205 ... we will take into account those comments because sometimes they think words
206 rather than score and they'll say look it tasted really something or other and you
207 think why didn't you put the score up there. So we'll often take into account, now
208 we don't actually manipulate the data, because you get fired from Establishment
209 1 for doing that but sometimes if, when we're interpreting the data, if enough
210 people have made this qualitative comment then we'll actually use this information
211 more than the actual numbers

212 ... You can't do that with the selection process the way we've got it designed at
213 the moment because we're not looking for numbers ability then

214 ... with copy sets we might find that a certain F ratio or significant score is what
215 we use as well

216 ... If they went a whole week with an acceptable level of standard deviation and
217 I'd given them a wide enough range of samples to cover most of the attributes I'd
218 be happy with them then and if 2 weeks prior to that they were all over the show
219 then I might even extend the half-blind period.

220 ... 3 full sessions of really good standard deviation on a wide range of products,
221 I'd be happy

222 ... I'd probably throw in a control every time, something that's just standard so
223 that you're actually testing their consistency cos you need to find out... if you
224 threw any duplicates in then they should actually score it the same. If they didn't
225 score it the same, then you'd be concerned so you'd always throw in one reference
226 every time and I'd throw in, for the standard attributes like sweetness, creaminess
227 or whatever, I always have something that is representing that every time. And
228 then I'd perhaps throw in one or two odd ones like a really high sour one or a
229 really high oxidised one or something like that. It will depend on what you've got
230 on hand but that's the procedure that I'd take. So I'm really looking for...I'd say
231 50% or more are standard type samples you know routine stuff cos that's mainly
232 what they would be doing and the rest of it just one off type stuff of any other
233 type of stuff they're likely to encounter... over the week

234 ... if you have three really bland samples, you know really high oxidised ones, you
235 are going to get a higher score than if you had put 3 or 4 oxidised ones together,
236 so you have to take into account contrast bias as well

237 ... normally I would try for something, say a reference and then sort of build it up
238 in intensity so you've got a little bit of oxidised, medium oxidised and high oxidised
239 or something, and they're all randomised anyway so they're not all going to get
240 them in the same order

241 ... if I'd gone 2 weeks and couldn't see any improvement I'd say scrap this, let's
242 go back

243 ... I'd say anywhere between 6 and 10 sessions that's when I'd make the decision.
244 Cos, yeah if they're not capable at that stage then they weren't ready to go in in
245 the first place. So your gut feel or your standard deviation calculation was wrong

246 ... Go back to round table training

247 ... I'd take them back to standards, back to all the basics, what you'd used as
248 training for sweetness, what you'd used for each for each of the attributes to get
249 that scale delineated again because at the end of the day, we should have
250 references, say on a 10 point scale, we should have a sweetness score of
251 0,2,4,6,8,10. And I'd go back and sort out which one was a problem area, which
252 ones they're having the most problems with, so I'd probably spend say 2 weeks on
253 references again, and then try throw some more project work at them again

254 ... We have changed group leaders because of that too

255 ... Her confidence in taking panels has gone down so I actually sit in on sessions
256 now and try and help and this buddying system that I've been trying to get going
257 with the other ones as well to ensure that if someone is sick, that the other person
258 has much more confidence going in there knowing what has gone on before

259 **Expert 2 excerpts - session 1 & 2**

1
2 ... we get people who respond to that ad and we get them in for what's called
3 screening sessions which normally occur over a week. They come to 5 sessions
4 where they will be asked to do a variety of tasks normally includes identification
5 of basic tastes and may include doing what we call threshold testing where they
6 have to taste a series of solutions until they can taste and identify the flavour, but
7 that isn't always done. Secondly it will include normally triangle tests which is
8 where they are given three samples, two are the same, one's different and they've
9 got to identify the odd sample. In addition we normally get them to do some
10 descriptive work where they have little vials which have got different kinds of food
11 in them and they've got to describe the smell or something like that and we also
12 do a couple of other things which are mostly for fun and also to fill in time so
13 they won't get too fatigued...

14 ... We also try to look at the person's personality, try and weed out any obvious
15 undesirables and so in the most recent lot of screening we actually had a small
16 interview where each person... just trying to find out what their motivation was...

17 ... Then at some stage we set a level above which we will accept people and below
18 which we won't...

19 ... We normally allow for more people than we actually want because we get, we
20 always get a percentage who are unreliable, not interested... a whole bunch of
21 things... and then from that the training starts. In my mind there are two different
22 things that we have to do because sometimes we screening for a total new product
23 that we haven't done before and in other cases, which is what I'm doing at the
24 moment, we have an existing panel which we need to add people to so I have to
25 try and train those people to fit in with the existing one and that's a little bit
26 different to training for something completely fresh and things...

27 ... If it's a completely fresh product which we haven't done before, don't have an
28 existing panel for, we normally would spend a few sessions, maybe 3 or 4 just
29 getting into word generation so you give them different examples of the product
30 and get them to just try and describe it and try and steer them away from
31 subjective words like bland or nice or yuck or whatever to actually trying to say
32 well you thinks it's nice but what about it that makes it nice or what about it
33 makes this thing disgusting. Things like that. At the same time in the introductory
34 sessions we go over things like how to taste, how to spit out nicely in public, yeah,
35 and a little bit about what the section is about and what the institute is about.
36 Need to try and give them a little bit of knowledge...

37 ... Trying to group the word together into broad categories and then ultimately
38 agreeing on the one or two words to describe that type of flavour. Following on
39 from that we start trying to develop reference samples for each of those areas so
40 like for sweetness you have to find something that's not sweet right through to the
41 limit of what we could expect that sort of thing, and hopefully some points in
42 between...

43 ... At some point along there we'd start introducing some kind of scale depending
44 on what we wanted to do with them. At present we're using what's called a
45 category scale, don't know if you're familiar with that term, like scoring something
46 from 0 to 10 is sort of like category scale so we might start them off by trying to
47 use words like slight, weak, moderate that sort of thing and then transfer them
48 onto numbers when they've had a bit of practice at trying to put it into categories.
49 Another kind of scale which we use for some panels is a line scale, 150mm long
50 with an anchor point at each end. So with that we'd start them off with that rather
51 than using words...

52 ... we're starting to try and find things which are reproducible and are reasonably
53 consistent for the panel...

54 ... Along the way, you should be trying to give them feedback on their individual
55 performance...

56 ... it's something that we're going to add onto, we do basically the same things but
57 we have to try and guide them a little bit so that we don't end up with 2 different
58 lists of words...

59 ... So we'd also be getting the existing panellists to join that group very early.
60 Basically as soon as they've got over their first couple of weeks, they've just got
61 their words and are trying to put them into groups so that we can make sure that
62 everyone understands the standards and if necessary we will use some of them....

63 KE

64 Mmm. It must be a lot harder to get reinforcements onto a panel than to build
65 a whole new panel... to get consistency.

66 ... A lot harder, because you tend to end up with factions.

67 KE

68 Yeah, the dynamics are disrupted. At what point do you declare a panel void and
69 start again for a product?

70 ... That almost happened for the chocolate panel because, for two reasons. One,
71 they got very low in numbers fairly sudden and secondly that we started to get a
72 really big variation in their data...

73 ... I didn't train the panel originally so it has been quite hard taking them over
74 because it seems to me that perhaps they didn't all agree on some of the
75 definitions or their understanding of them was not clear...

76 ... normally it's a matter of going back and trying to fix that before going on and
77 doing somebody's project...

78 KE

79 How many is not enough for a panel?

80 ... Less than 8...

81 ... we start off with about 20 and they sort of whittle down within the first week
82 or so to about 16 or so. We tend to lose a few. Around 20 to 15 is a good
83 number. We can then get them... if they're going away on holiday we don't drop
84 too low. If you only have about 10 and one gets sick, you can really have a
85 problem. Yeah, around 15 is a good number. 20 is hard to control...

86 ... In theory, we declare it trained when they can reproduce or at least evaluate
87 the same samples, when all the standard deviations are within the limits that we've
88 set and sort of remain there quite happily. A third factor that comes in quite a lot
89 is the pressure to have them ready...

90 ... The group leader. I mean I would discuss it with Expert 1 but I'm sure what she
91 would say is, if you don't think they're ready, don't do it. We might decide that
92 we'll do a round table discussion but we won't do any statistics or anything if
93 there's real pressure on. That normally is the group leader's decision...

94 ... went back through whole heaps of data and worked out what the average
95 standard deviation was for the different attributes and then we tried to set limits
96 that we thought were reasonable and achievable but not so easy to reach that they
97 were nonsense. Most products or most attributes it's plus or minus one on the
98 scale but some of them go up to 1.5, for attributes that we know the panel have
99 more trouble with. So yeah, you've only got a 10 point scale so when the standard
00 deviation starts to get too high it sort of makes a nonsense of it, of the data.

01 ... you want to show that the panel is able to score using the same part of the
02 scale and can do it consistently...

03 ... a really simple product I would say 8 weeks is really the shortest, that's with
04 them coming to 3 sessions a week. And the longest... oh 6 months maybe...

05 ... things like baked products. I anticipate that chocolate is going to take quite a
06 few months. 3 is sort of about the average. Things like butter and whole milk
07 powder normally run for about 3 months. You tend to take longer with a product
08 where the flavours are more sort of tied up together, and it's really hard to break
09 them apart and treat them individually. Things like yoghurt. The flavour is
10 reasonably simple... some of the things like chocolate and baked products it might
11 be several weeks before we get those samples and we've got to go back to what
12 they were talking about 2 weeks ago. The chances are they will have changed their
13 minds anyway, it's a much more difficult process...

14 ... Something like chocolate, we've got what we've got and if we want to try
15 something out it might be, you know, by the time it has been manufactured it
16 takes 3 weeks before people taste it to for it to set up properly and it might be
17 3 weeks before there's time available to make that sample. So any product that's

liquid we can prepare here, and that really helps the training process because you can listen to what the panel's stating one day and say tomorrow we'll try x, y and z, and do it...

... time that it comes out is at our performance review, like if I took a whole year say to train a chocolate panel then that would come out in my performance review. Okay while I might have been doing everything that I could think of it probably would be that I should have been seeking more help or getting more support or... admitting defeat. I don't know, because to achieve at least half of the objectives that we have your panels need to be performing and if they're not then those objectives will suffer and that comes back in that form...

... normally 3 sessions a week and they last anywhere from... they can last up to an hour...

... a big group they take longer and like, for a panel that is trained like the butter panel they still have about one training session a week or a fortnight, depending on their schedule. Some of them are quite quick, 15 or 20 minutes because there isn't that much... it's more for them so they can see how they score compared to everyone else. I would say that the average is about 40 minutes...

... The main reason as far as I can see that we have them on regular times and days and things is so that they get into a routine...

... if you only did 2 sessions a week you didn't actually achieve that much because you might start of on say Tuesday working on caramel flavour and it might take 3 sessions or more to work through that. If you only have 2 then a big long gap, you will have to repeat what you did on the last day. 4 we decided was asking a bit too much of people. we didn't there was enough dedication in the community to get that, so yeah, 3 was a trial thing but it works really well...

... We pay them \$10 per session regardless of the length of the session so we do 10 minutes or an hour, they get \$10. They get paid by cheque fortnightly. Well technically they're not paid, it's reimbursement so they don't have to pay tax on it. Legally they're volunteers. Also with some products we bribe them with food. Like the butter panel might get chocolate afterward because when they first started they all moaned terribly and were really offended that they would have to eat butter and other people got to eat cake and chocolate...

... They get paid right from the time they come to screening. They get paid less for coming to screening. Normally if they came to a 5 day screening they get \$30. And they get paid right through their training, that sort of thing. We're also starting to introduce some sort of bonus scheme, not on a regular basis but just every few months for those people that have attended more than say 90% then they'll get a small bonus like a book voucher or a petrol voucher or something...

.56 ... for the initial setup of a panel, you can have multiple clients and Establishment
.57 1 in fact will be paying for that, that'll be as part of the working expenses of the
.58 unit but then if we go training for a specific project then that's project work...

.59 ... I'm responsible for 2 panels, one of which is trained and the other is in
.60 training...

.61 ... One of the things that we do in training is put all the scores up on the board
.62 and everyone can see what everyone else has put. Someone who is a non-
.63 performer tends to show up then and normally I would talk to them and say, look
.64 KE you put 2 for this item and everyone else has put 10, perhaps you'd like to go
.65 back and try again and see if you can understand what they're talking about or if
.66 there's some other flavour or some other way you'd describe it. It's very difficult
.67 though because often when you do that, people go back and say, yes I can get it
.68 now and you really don't know whether they can or not...

.69 ... we've done various projects for people and the results have been confounded
.70 by the variability of the panel and it looks like there's something there but there's
.71 so much variability in the data that you can't draw that conclusion confidently, and
.72 so you've got a client saying, look, we're paying you x thousand dollars to do this
.73 and you can't tell me because your panel is so variable that you're not sure. And
.74 that reflects badly on me and on the section and on sensory as a whole and so to
.75 improve that variation within the panel has to be dealt with...

.76 KE

.77 When you've got a person that's performing well, you trained them up and then
.78 performance starts to decline, what causes it?

.79 ... Probably depending on the individual. Some things are that they get bored, on
.80 a tasting, they sit down, write down a few numbers, take the money and go home
.81 again sort of attitude. Maybe not enough on-going training like if I put them in
.82 the booths week after week and don't give them any feedback. People may start
.83 going off on their own tangent. Stress, pregnancy... got one of those at the
.84 moment...

.85 ... Every panellist has their bad days or bad day or two...

.86 ... panellists have actually commented that... see often a lot of the samples are
.87 really similar... and after maybe the second week of it they sort of start to think,
.88 gosh there must be something wrong with me, these all taste the same, you know,
.89 I must be missing something, and they start getting really paranoid...

.90 KE

.91 What do you do with people who are good performers but through fatigue or
.92 adaptation they're just losing their sensitivity to a certain product?

.93 ... I guess they're probably quite difficult to identify

94 ... making sure that they're remembering what those reference samples are and
95 they're scoring things according to those samples. Panellists have a real tendency
96 to want to score samples relative to one another on the day and so therefore you
97 can never get day to day consistency. So yeah, going back over those reference
98 samples regularly, I think can help stop people becoming desensitised but you also
99 get the opposite effect that they can get over sensitised and tell you that they can
100 taste a hundred and one different flavours in the sample and maybe they are or
101 aren't but no-one else can taste them or you know... yeah, I think reference
102 samples are about one of the best things we can do...

103 ... we still limit it to the three or maybe four sessions per week and for most
104 products we limit it to about four samples per session...

105 ... it starts when we have enough requests programmed waiting for a certain
106 product type to justify setting up a panel and the decision is made to do that and
107 someone has to be responsible for it...

108 ... you're responsible for producing that data or having the data produced and
109 reporting it and being accountable for it, in that you can understand, explain and
110 justify what was said about it...

11 KE

12 So is the screening generic or is it related to the product?

13 ... It's a mixture. Like a generic base but with things that can be added on or
14 taken off depending on what it is we want to achieve...

15 ... in some instances we just know that say there's going to be heaps of cheese
16 work and we're going to set up a panel and the project information comes
17 afterward or during the training process and other cases we know before we start
18 like Technician 2 and Expert 4 are the classic examples. They know exactly what
19 it is that the panel is required to do...

20 ... you have to sort out times of the day when you're going to run your screening,
21 how many people you want to attend, how long it's going to take them, the
22 equipment that;s going to be required and the samples and everything required.
23 Any publicity information you want to give them. Who's going to do what tasks
24 on the day...

25 ... During screening we have 2 approaches. One is that you can plan it all out and
26 say right this is what we're doing on days 1 through to 5 end of story. Or if you're
27 about to do something that you don't have much experience in, what we might do
28 is say plan the first two days and go through the results and say well the screening
29 is far too hard. No-one can pick the differences between samples, it's too difficult.
30 By making it too difficult you're not going to achieve anything, because you're not
31 going to select anyone. So we'd normally want to try and find a level which would
32 spread people out and show their different abilities to discriminate Ok. So you've
33 got to try and find that level fairly quickly, because we only have 5 days to do it.

34 We have a lot of base knowledge with which to do that, so it's setting up the
35 actual screening...

36 KE

37 You mentioned earlier on that you interviewed you potential panellists. Who does
38 that?

39 ... The panel leader does.

40 ... I had a look at their data and sort of pencil in those people that I thought had
41 potential and so trying to take particular note of them and to see what they'd be
42 like...

43 ... Interviewing is an aspect of selection....

44 ... group co-ordinator. Actually standing up there and trying to see the wood for
45 the trees. The other one is more of a physical task I guess like organising the
46 samples, co-ordinating them, trying to select samples which might help the group
47 in what we're trying to say. At the same time reading the literature or past reports
48 or whatever and trying to understand that from a food science point of view but
49 also trying to be a facilitator, you know, without trying to put the answers into
50 their head...

51 ... One part of it is the researching side, trying to understand any reactions that
52 might be going on in the product, reacting compounds and that sort of thing. The
53 other one is maybe a little bit more on the technical side and may actually be
54 delegated to some one else, in that it's the physical handling of samples, checking
55 out sample availability, preparing sheets fro training because often during training
56 we might go through half a dozen sheets or more...

57 ... There's the analysis and interpretation of the data that's being generated during
58 training..

59 KE

60 Right. They're all interrelated because otherwise it wouldn't be one person's job.
61 Could we say that analyst is a task?

62 ... Yep...

63 ... General communication, like communicating with Expert 1, communicating with
64 clients, communicating to the other staff here as required...

65 EXCERPTS - SESSION 2

66 ... giving the panel confidence and showing them that they are doing well and that
67 their scores are meeting whatever criteria we've set, whether it's standard
68 deviation or range or whatever we've chosen...

269 ... giving them feedback sort of like that day or the next day on how well they did
270 and if we repeat that part, show them what their scores were over certain days or
271 whatever, just so they feel good about themselves...

272 ... be confident in myself that each of the panellists do know what the attributes
273 are and that they can score them reasonably well and that should have been
274 achieved by the use of standard samples and things...

275 ... start bringing in project samples so if cake is going to be something unfamiliar
276 which they haven't come across before, which quite often happens, or it's like
277 what they've had before but because it's not a standard sample which we've
278 manipulated, it's not a naturally occurring flavour or whatever. It might be in the
279 same flavour group but not exactly what they're used to. They have to learn to
280 recognise that and know where to put it, rather than sticking everything in the
281 'other' category because it's not exactly as they've learned it...

282 ... don't normally tell them too much about the aim of the work or anything like
283 that because it can result in quite a lot of bias if you do that too much . It's
284 difficult but it's good to try and find some sort of balance so they know a little bit
285 about what they're doing but not so much that they then skew the data for you...

286 KE

287 That's good. In relation to you actually confirming whether they understand the
288 attributes and can scale them or not, what actual range of tests do you put them
289 through to do this?

290 ... Traditionally, it's just a round table training session like in here and yeah, giving
291 them reference samples and making sure people can recognise them... like in
292 training they might be doing say a flavour which we call 'cow-ey' and they might
293 have had it at three different levels, OK and they get those often in a similar
294 order and if you suddenly throw one in amongst lots of other samples they still
295 should be able to recognise that...

296 ... The other thing to do is to put them in the booths and make them do it by
297 themselves because a panel that does all their work in here for 3 months and all
298 of a sudden they have to do it without any visual feedback, they often just fall
299 apart for a couple of days...

00 ... in every panel, there's always at least one attribute which I'm a bit
01 uncomfortable with cos I think the panel flounders a bit. So I might give them
02 more leniency on that

03 ... at the end of the training, I'd expect only to have to do it a few times because
04 their training should have been good enough, but I have confidence in that and
05 it's just a final check. It isn't something that we do that formally, just the last few
06 sessions we take particular notice and maybe give them some of their data back
07 and that sort of thing. Point out to the people if they're consistently scoring lower
08 than everyone else...

... at the end of it, they should be able to focus on every attribute in the sample, saying is it there, isn't it, how much...

... When I looked at the data, I'd look at every attribute, but the sample that I gave them might have focused on one or two...

... depends on the panel, and the attribute and the product. Normally we like to have the standard deviation round plus or minus 1, that's on a 10 point scale. But like if we have an attribute which we know is the weaker points of the panel then we can allow but 1.5 is pretty much the end of the line. Getting above that makes your data pretty fuzzy...

KE

Could we just go back to the testing phase, the bit at the end of the training cycle. What happens if you have the gut feel that this is the right time for them to finish, or there's a bit of commercial pressure from a project sitting there, and the performance simply isn't there, like the standard deviations are beyond 1.5.

... Sometimes the work goes ahead anyway but normally we tell the person, look, we can do it but the panel, you know, isn't up to scratch but if that's what you really want to pay for then, you know... but normally, no...

... if they had standard deviations of 3 or 4, then you would have to look at are the reference samples inappropriate or maybe you've got some panellists that just aren't up to the task. I mean, when it gets to a situation like that, you really have to start looking hard at individuals. Is the standard deviation high because everyone is putting 2 or 3 and one person's putting 8. Or is it because the panel is completely spread over a wider range...

... If they are consistently causing the performance of the panel to be low, then they need to be told that, and if it's something that they can't help then they either have to learn to adjust or else they need to be off the panel...

KE

So do you keep individual records of the panellists to track their performance?

... No, we don't. Sometimes, like for a specific project, I might do an analysis and look at individuals but it's not something that we do standardly although I hope that when compusense is installed, we will be able to do that a lot more easily...

... The first step is to do more training on it and to try and find out from the panellists if they think that they really understand what that attribute is because particularly here I think, there's a real tendency at the start for the panellists to want to please you and so they say, you give them a sample, and you say we thought the sample might have whatever characteristic and they go yes, yes, yes and if you think yay, I've made a breakthrough and if they haven't been completely honest with you, then you might think that they understand that standard and everything's home and hosed

348 ... it has caused problems with the chocolate panel because they get an unknown
349 sample like in project work, they suddenly don't know what it is and they don't
350 know how to score it because they've learned the process the wrong way round.
351 You have to ask the question, is it because somehow you've tried to impose a
352 word on them that they're not comfortable with. Is it that the references aren't
353 appropriate. Is the attribute even really relevant to the product?

354 ... you have to go back and look at what you used as training standards, whether
355 with hindsight something else could have been more appropriate, whether in the
356 experimental product it's maybe a sub-group of that flavour that's coming out or
357 whether it has gone beyond that flavour into something new

358 ... If it's one or two individuals then you have to look at whether they have at the
359 same rate as the other people or as well or have the same understanding, are they
360 really understanding the scale or are they even understanding what's being asked
361 of them. And if not then you have to make a choice as to trying to fix it or to take
362 them off the panel

363 ... how many people have I got rid of? only about 2 off one panel. Expert 3's got
364 rid of... sounds terrible... taken off the panel I think 1 person off the baked
365 products panel

366 ... if they have some sort of chemical blind spot in which they're just not capable
367 of perceiving the chemical that's causing that flavour

368 ... Another quite likely cause is that, for example in chocolate, we measure burnt
369 flavour and there's always been one or two people on the panel that the first time
370 we give them a burnt sample they say it's nutty, OK. Don't ask me why but it has
371 happened. And if I as a panel leader or that person fails to recognise that
372 connection, then that person will carry on scoring that incorrectly

373 ... a lot of it's panel leader error or interpretation, and yeah, I would say that one
374 of the most likely causes is through the training

375 ... The panel leader has to be really tuned in to that kind of thing in that if you
376 have a panellist who is consistently not perceiving something, instead of just
377 writing him off as useless, trying to see, is there something else that's going on

378 ... I can't think of any formal checks that we have, on how the panel leader has
379 done other than how it's reflected in the data and group. And that doesn't mean
380 to say that if the data isn't that good that the panel leader has necessarily done
381 a bad job

382 ... that person has a responsibility then to ask for more help or get more support
383 from within the section

384 ... most panels we have tried to have like the panel leader and someone else in
385 the section who knows a reasonable amount about what's going on with that panel

386 ... a panel is never perfect and I don't think I'd say that they're never trained but
387 I'd say there's always room for improvement so the work I'm doing at the
388 moment, I'm making sure that their data is being collated every day after the
389 panel and I'm putting the mean and standard deviations on a graph out for the
390 next panel so if I need to I can repeat that day's tasting or I can talk to individuals
391 who may be having a problem. or if the whole panel has gone off the edge then
392 we can stop it, have a few day's training, talk about the problems and go back and
393 start again. When it happens during project, the problem arises if you're not on
394 top of it and that's something that we haven't done that well in the past but we're
395 starting to be a lot better at it because if you let the data mount up for three
396 weeks while you're doing project work and then find that half the days are
397 absolute garbage, I mean OK you could go back and start again but a lot more
398 extra work and a lot more expense so it's, yeah, you're much more in tune with
399 what's going on if you're looking at it daily

400 ... compusense measures it in pixels and it's limited to 61 pixels, whereas before
401 we had a 150 mm line, so we have reduced the sensitivity of the scale by more
402 than half and that's quite a concern to me. it might make my standard deviations
403 look better but I also think that for the panellist that it's much harder for them
404 to get the sensitivity that they feel they've got on the scale

405 ... Within 2 or 3 minutes after the panels finished you can get mean and standard
406 deviation, range and you can get them to wait in here and show them

407 ... if you wanted to make it a very clear cut decision then we need to have set
408 standard deviations and perhaps ranges, some kind of defined tool to measure
409 them by

410 ... we don't have it formally defined how many days in a row they should be able
411 to do this or like with duplicate samples, how close they should be, like within a
412 day or over a day so yeah it probably needs a little bit more definition if you
413 wanted to be very strict about it. But also it's difficult because there's also a
414 subjective side to it like if it's just 1 person that's messed up, does the whole panel
415 suffer because of that? and if they don't do it again for a week and then it's
416 another person, do you keep holding them back because of that so it's almost like
417 you need a band with a critical level on either side

418 ... when they finish training and begin project work, tend not to tell them because
419 it creates this great excitement and they're all expecting... different samples that
420 often are very standard and very routine and so there's a psychological thing

421 **END OF EXCERPTS - EXPERT 2**

122 **Expert 3 excerpts - session 1 & 2**

1
2 ... we advertise and as a result of that we get 60 to 80 people and you set them
3 tasks. Well, when I'm running the screening I set them tasks that are related to
4 the product that they'll be evaluating. We go through, first of all, the basic tastes
5 and go through some difference testing and then some difference testing with the
6 product itself using project variables to make different samples and that's
7 supposed to get harder as the work progresses and they get more used to it but,
8 yeah, it's quite difficult

9 ... I'd do a bit of personality work as well, just group dynamics

10 ... And then training starts and you just basically do an introduction to get them
11 all familiar with each other, you know where Susan comes from and all that so
12 they can get down to what they're doing then

13 ... just vocabulary development, that's probably the first stage where they generate
14 lots of words to describe lots and lots of different samples and you try to get
15 samples as varied as possible at that stage, so that they can really get any sort of
16 attributes that they can think of. Then later on, if you're just doing flavour or just
17 doing texture you start to concentrate on those. We probably use the same
18 samples to begin with over again. Then introduce a preliminary, basic scale like
19 absent, weak, moderate, high and try and pick out individual attributes and work
20 on those

21 ... So if they all agree that their milky is say similar to someone else's creamy or
22 they all agree that it is present in that attribute then I'll put to the panel that I
23 think that everyone is calling the same attribute different words so what can we
24 call it that everyone will recognise. And you go through that and that's a really
25 circular thing

26 ... then you start to work on scaling each attribute with that basic, preliminary
27 scale and then trying to get it to be a bit larger but at that stage a few decisions
28 have to be made by me regard what type of scale to be used

29 ... then setting up standards along the scale I suppose is the next stage. So the aim
30 of it is to have them all understand each attribute, understand the scale and be
31 able to recognise specific samples along that scale for that particular attribute

32 ... put them into the booths and do a few tests to see how consistent they are

33 ... They've just shown interest from filling out a questionnaire that we have, they
34 go on to our database and then, yeah, if they're available at the time of our
35 screening and they're going to be available for the next however long then they're
36 screened then based on their performance in the screening test they're put on to
37 the panel. That's if they're 75% correct... On the tests that they've performed for
38 the week's screening. In the two panels that I've done they've had two weeks of
39 screening

40 ... like they might get 8 samples and they have to taste all 8 and put them into
41 order of increasing sweetness. That's one sort of test they can do or threshold test
42 where they've got a sequence of 8 and they have to say when they can taste a
43 difference from the reference... just a detection threshold really, that there is a
44 difference and then a recognition when they recognise what the actual taste is,
45 that's one other test. Difference testing is used which is, well there's several
46 different ones but the common one that we've used is triangle tests which is with
47 3 samples and you just circle the sample that's different to the other two. The
48 description tests we do with a sample and they have to describe the aroma

49 ... The last 2 times they did interview them. I don't know what they asked them
50 or anything like that. I haven't interviewed any panellists

51 ... I think I'd run a training session like with bizarre samples, different sample you
52 know and just put the group together just to see who really takes over or who
53 didn't

54 ... I think it's quite important to have people that are really neutral, who are
55 willing to learn

56 KE

57 Mmm. So how long would you have them and for how long in the selection
58 process?

59 ... Every day for 5 days... For about an hour. Yeah, but because there's like 60 to
60 80 people it would be a whole morning. You'd have groups of people coming in
61 and going all the time

62 ... I'd say 15 to 20 is what I'd choose cos there's always quite a rapid drop out rate
63 in that first 3 months, people that don't like it or don't want to make the
64 commitment, that sort of thing so yeah, and then I'd expect that to drop down to
65 about 12 to 14

66 ... To run a panel you have to have 8 panellists... You can't work it out
67 parametrically if you've got less than 8

68 ... Well they'd come in for the preliminary first week and basically the first day is...
69 well probably the first week, we really introduce them to what they have to do and
70 we really push the things like don't wear perfume, don't clean your teeth before
71 you come, you know, you're just reinforcing the principles like just do your own
72 thoughts and all those things. And also start looking at any varied samples. And
73 on the first day they introduce themselves to each other and give a background
74 and I just then start them off writing down as many words as come to mind about
75 each sample that's in front of them

76 ... I'll tell them what each sample was and talk about it a bit and then probably
77 the next week, start looking at the words that they've written up there and ask
78 them if they can see any words which could be similar. At that stage you just start

79 varying all the samples that you can think of and if there's one word which keeps
80 coming out a lot, you stick that down on the form, so they've got that to use... It's
81 at that stage that I get caught up in trying to produce samples that are going to
82 be accurate enough... Accurate enough to get everyone to agree to a particular
83 attribute...

84 ... the more complex the product the more horrendously difficult it is to get
85 accurate attributes

86 ... we have got just a general thing on sensory evaluation and it has got things like
87 the tongue, you know, where the stronger areas on your tongue are, not to brush
88 your teeth and that sort of thing. I can't remember everything that's on that
89 handout. It's just really general stuff but on the back of it is a list of things...

90 ... I'll introduce a scale like probably the first scale that I would use would be
91 absent, weak, moderate and strong so they've only got the A, W, M or S for each
92 attribute

93 ... our preliminary evaluation form?... it has got the list of the attributes, it has got
94 the scale and they just have to put the capital letters that they think are in there

95 ... I try and get people to interact as to how they're tasting things...

96 ... It doesn't actually last that long with that sort of scale. We just try and get them
97 to go through with each of the attributes, then I normally introduce them to a
98 category scale in any of the work that I've done. The primary reason for that is
99 that historically that's what Establishment 1 has used

00 ... it's a 0 to 9 point scale and they can only use the points, it's not a line scale,
01 and the distance between points is not set so the points are anchored, like 1 is the
02 threshold level, 3 is weak, 5 is moderate, 7 is strong and 9 is intense but there's
03 no actual linear, no mathematical distance between the points

04 ... basically what I do is I go through different samples, getting them to practice
05 on the scale and just discussion until they've got a good grip of what they're
06 doing... I'm analyzing the results throughout the time and giving them feedback
07 about how they're going. So basically that's it until they're competent at each
08 attribute and we've got a system where, I don't know if Expert 1 showed it to you,
09 where we've calculated the standard deviation about the mean for the panel. If
10 they're too high, if there's too much deviation, they're not ready

11 ... we redo that work where the standard deviation is too high

12 KE

13 Right, and what is the formula? or is there a set of data that you're comparing
14 against?

15 ... Just whatever seems realistic

16 KE

17 How long then does it take you to get a trained panel?

18 ... 6 to 8 months I think. Milk fat took 6 to 8 months. I mean the literature says
19 2 to 3 months

20 ... Feedback of their results but also getting real work. They love seeing their own
21 progress. They love seeing graphs of themselves and what the average was and
22 what the standard deviation was and who was too high and who was too low

23 ... if they're consistent, regularly turning up, we give them money and things like
24 that

25 KE

26 Yeah. What causes performance to drop away other than lack of motivation?

27 ... Lack of variation in the samples

28 EXCERPTS - SESSION 2

29 ... Well I start off, I just have a vague idea of what sort of product and the
30 particular attributes so I start off and I just get really different, like if I can get
31 samples from overseas that are really different

32 ... I'll get them to profile things and just profile standard samples that they'll
33 probably be getting in project work and then I'll probably, in their results I'll just
34 have a glance at their results and see that there is no-one right out of the ballpark
35 but probably by that stage that's not likely because they've been weeded out to
36 a certain extent

37 ... I'll do an ANOVA between panellists, and then I'll look at the mean and the
38 standard deviation and if the standard deviation is quite high, then they still need
39 training because there's too much noise around the mean. If there's one person
40 who's outlying on a particular attribute then I'll get them, you know, work with
41 them on that attribute and do a training that's specifically geared up for that
42 attribute so I can see how they perform. I give them some stats analysis like
43 principle component analysis but I've found that lots of those multi-variate
44 analysis methods are too generalised and it's just looking at their consistency
45 really. If I do the same sample twice in a row I see whether the same panellists
46 have given them the same scores

47 ... Over 2 days but that's probably what I'd do and if that all checked out. There's
48 no formal sort of stages that they go through but if I get a feel that I've got
49 confidence in them, basically they'll do

50 ... We'd probably do each attribute 2 times

51 ... low standard deviation and consistency in the attributes.

... if I were to change from a 9 point to a 10 point scale and the acceptable standard deviation was 1 in 9, then I'd just take that percentage and apply it to the 10 point scale so that would work out to be 1.2 or something

... If I had the training and they're all over the place and they don't understand the samples then I'd say come back tomorrow and I'll see if I can get more... like we have the training and they might not go very well on one particular sample and it's quite an important sample or something and I'll think I'll get them back next week and we'll still have this sample perhaps with a different angle, so that's what I'd do first and if they still can't get it I'd just take them back to basics again and start them on a really wide variety of samples and then try and tune it down again

... then 6 weeks down the track after you've done 2 more blocks you might go back to the block that you did 3 times ago and see how they're still going with that, that's more realistic

... I don't tell the panel what to taste sort of thing. If they're saying to me, we just don't understand it then I sort of drop it more than anything else

... drop the whole tack that I'm on at that time, try and change tack and maybe try and go around the particular attribute that I'm working on

... I just basically kept training them and training them and changing tack and repeating things trying to get things that can taste and stuff, to understand it

... I'd put them in a panel just to see how they'd go and I've probably trained them with the samples that they'll be evaluating anyway and if they just way out of the ballpark on them then I'd put it off probably. But yeah, that has its consequences but rather than get results that are quite fuzzy

... If I had a panel that was consistently not performing I'd go to Expert 1 and I'd explain to her. What would Expert 1 do? Expert 1 would probably ask me if I'd done a lot of things, suggestions to try and get them back into line and I'd probably go away and... I'd to be able to make the decision that they're worth it or they're not worth it but I'd have to make that decision with Expert 1 based on my recommendations I suppose.

... I try to get real samples like I try to get things that have been rancid for 8 months, and then something that's moderately oxidised and put them in the booths and see whether they still say that they're moderately oxidised. It's sort of not doctored samples because my interpretation of doctoring might be wrong for that sort of thing so I try to make them as real as possible

... Probably a week, I'd say a week of panels going into the booths, coming out and training straight after that and probably be a week with the samples that they'll probably be using in the project plus some of the old references that they'll be familiar with

90 ... They'd probably be duplicated so that I could see if they were consistent within
91 each person and then I'd take them back into the round table room and make
92 sure that everyone recognised that the lactic acid, well the different flavours that
93 were in the lactic acid were more or less standard

94 ... what I've tried to do is suss out the parameters of shortbread flavour, that's
95 what we've been doing. I've worked with the flavour chemist, Russell, and he's
96 been trying to enhance specific attributes within the shortbread and I think that
97 they're just starting to understand specific attributes

98 KE

99 Right. How many times would you accept them doing... if the standard deviations
00 are all good across the group for a certain attribute, how many times do they have
01 to produce that result before you can say, OK, they understand what butterscotch
02 flavour is?

03 Some of them, in excess of 3 times.

04 ... on one particular attribute it goes from 0 to 8, I'll say to the people who put
05 8, why have you put it so high, please explain to everyone else. They'll explain it
06 the way they thought it, and then I'll say to the person who put 0, now can you
07 relate to anything that person said and they'll say yes or no. I'll say why have you
08 got 0, what was it about the sample that made you think that there was none of
09 that particular attribute. If they can't agree then I'll ask everyone else if they can
10 understand what's going on and if they can't understand it... I mean quite often
11 when it's 0 to 8 it means they just don't understand what the attribute is

12 ... I used to have a cooked butter and an uncooked butter flavour in shortbread
13 and uncooked butter was consistently all over the place. They'd say they could
14 taste uncooked butter and when I got them to talk about it they didn't know what
15 they were talking about, they were just talking through a hole in their head so i
16 just took it off. Probably what I would do is just take it off. I said to them, if I can
17 find a sample that typifies this, then we'll put it back on but unit I have a sample
18 that you guys can consistently find uncooked butter flavour, then I'm not putting
19 it on

20 KE

21 In your experience, how long do they tend to go on for, from the time that you
22 sort of get a gut feel that they're about ready, how long does it go on from that
23 point to you saying they are trained?

24 ... I take ages to train my panels. Probably 6 weeks

25 ... I told them there's none there, because everyone else has said, they don't
26 understand that so they've got to understand each attribute and each scale which
27 I think they need to learn themselves

Appendix D

Excerpts by Meaningful Phrase

228 **Expert 1: excerpts by meaningful phrase**

1 We get them straight off the street.

2 We select them on the basis of how good they are at tasting a particular product
3 and in order to do this we employ threshold testing
4 which is basically how good they are at the four basic tastes.

5 We need to find out how good they are at discriminating between certain products
6 for example a cheese panel gets to try all cheese products.

7 We also use triangle tests where 2 samples are the same and 1 is different.

8 They need to be able to spot the difference.

9 At the end of 1 to 2 weeks we eliminate about 75% of people.

10 That usually leaves about 15 to 20 people.

11 Now some of these people are good at one product but not at another.

12 Then in the first few training sessions after selection
13 we get them to describe everything about the product
14 for instance flavour, texture, colour.

15 We take all the words that are used
16 and the group leader finds matching words to get a shared vocabulary.

17 That usually takes a couple of days.

18 Then we introduce intensities.

19 The group leader works out some end-of-scale words like weak, slight, strong,
20 and then moves from those onto numbers.

21 We would probably spend 2 to 3 weeks on words.

22 In actual fact, the process of converting words to numbers is easy
23 by tying values to specific words.

24 Then the pressure goes on the panel
25 because this is when the leader sorts out the outliers.
26 We can spend 1 to 2 months getting the numbers right.
27 We also start measuring variation,
28 analyzing the standard deviations
29 and giving that feedback to the panellist.
30 At the end of this process we make the decision
31 as to whether this panel is fully trained or not
32 Once a panel is trained, they are 100 times more sensitive
33 than a person on the street to that product.
34 We've got to try and get rid of the variation
35 between the experts at that decision point as it's critical to the quality of the panel

36 KE

37 Have the group leaders been given feedback about their variation in the same way
38 that the panellists do?

39 No... there are definite differences in assertiveness
40 and how hard each one pushes a panel.

41 There are standard deviations for each product based on historical data
42 but leaders aren't using them

43 so my question is what are they basing their decisions on?

44 You see the length of panel training ranges at present from 4 to 6 months,
45 depending on how motivated and keen they are.

46 It also depends on how good the leader's standards are and how pushy they get

47 KE

48 Right. So how many hours training per week do they do?

49 3 times per week.

50 In training that is a 45 minute session ranging up to 1 hour

51 During screening they get \$5 per day plus a \$5 bonus to be there at the end.

52 Normal pay for training is \$10 per session

53 regardless of the length of time that they're there

54 KE

55 Mmm. What causes the decay in performance?

56 Changing lifestyle, pregnancy, stress and tiredness.

57 Fatigue and adaptation due to over testing...

58 When we give results back, we hide the names

59 except for those who are consistent outliers.

50 What I want from this is something neutral and standard

51 for group leaders to refer to

52 It all starts I guess with interviewing.

53 The first interview that you're ever likely to do is going to be on the phone

54 and there's normally a standard sheet that you go through,

55 or if they don't want to do it on the phone, you send it out.

56 So the basic information that you've got on a panellist is what's on the sheet

67 like their sex,

68 what they like or dislike in food,

69 whether they're allergic to stuff,

70 how committed they are.

71 When they come in for screening, toward the end of the week

72 we actually interview everyone

73 and we ask them, from another standard sheet,

74 what their commitment is, are they going to get employed,

75 how happy they feel about coming in.

76 What we're trying to ascertain is if they are going to be a good group person,

77 whether they're going to leave us in the lurch 2 months down the track

78 and whether or not they're actually interested in it.

79 We try to find out without being too subtle whether they're doing it for the money

30 or whether they're actually genuinely interested...

31 I try to have 2 or 3 people in our section sit in on the interviews.

32 ... sifting through and sorting out the ones to be included.

33 I guess you could call that the personnel selector.

34 It overlaps with the interview but is just a confirmation.

35 Have we got a good group person in there?

36 But the main selection criteria is still the results from their screening tests.

37 It's based on the most objective information that we can get

38 so in combination with their results from the screening tests

39 and the limits set on them, plus the interview,

40 that's how we would select.

91 We tend to select more people than we need because they usually drop out.

92 If we were to select just on the basis of results, we wouldn't get enough.

93 We often drop the standard to increase our numbers.

94 ... You've got to take a whole group of strangers,

95 people that don't know each other and we've got to convert them into a group

96 of people who know each other and trust each other

97 enough to say well I really don't think that it is such and such

98 and feel comfortable with that.

99 So this is where your skills are really brought to bear as a facilitator

00 ... The first few sessions are really important ones as they set the scene

01 so we've tried to do some fun things and some interesting things.

02 Especially introducing two groups of people like where we are reinforcing a panel

03 and you have to bring the new lot and the old lot together...

04 we've done a few little things like blindfolded a new one

05 and the old one feeds them a product and they have to say what they think it is.

06 Then they whip off the scarf and introduce each other.

07 You have to be able to read people and know when to push and when not to.

08 One of the harder roles is knowing when to push them.

09 I guess that leads into the next task which is as a trainer.

10 ... you've got to analyze the performance,

11 looking at questions like why is the person doing things wrong,

12 what can I do to improve them,

13 what steps do I need to take to ensure that they do pull up,

l14 and if they don't making the decision to get rid of them.

l15 It's analysis I guess.

.16 ... It also comes in when you want to throw a bit of stats around the place.

.17 You want to try and link something that's happening with the panel

.18 with something chemical or physiological or whatever

.19 but analysis comes further down the track.

20 It's when things get tough that you try and get a lot more discriminating

21 and analyze a lot more

22 ... the panel leader has to make a final decision about the panel or the panellists.

23 We say that the panellists have got this apprentice period.

24 Now I could be really hard nosed and say right,

25 if you're not performing in 3 months, then you're down the road

26 but lots of other issues come in and affect that.

27 Like we have to keep persevering with useless ones to keep the numbers up

28 or they are going to bad mouth us if they leave

29 ... There's also decisions like are they ready now or are they not.

30 Other issues that affect that are if the project work is coming,

31 or getting nagged to make some money out of the panel

32 KE

33 Can a person that is socially disruptive bias the tasting?

34 ... Definitely, both in the results of the panel overall,

35 and also in the motivation of the other panellists

36 KE

137 I see. What I'd like you to consider now are some of the key items
138 that make up the tasks we have discussed.

139 If we can look first at the task of interviewer.

140 What do you feel are the elements that make up this task?

141 ... the most important ones would be... articulation,

142 the ability to speak well to the applicants.

143 Social skills are also important

144 and we also have a checklist that we work to

145 I've got used to three or four months

146 so after they've been going sort of two or three months

147 you used to start looking

148 quite carefully at their individual results, ok.

149 So we would actually start using a calculator

150 or putting it through the computer

151 often we had at the time and getting out the mean

152 and getting out the standard deviations

153 and having look to see if that standard deviation is coming down

154 just before we consider them fully trained

155 we do what we call half blind tests,

156 where we take them through the training situation,

157 put them in the booth just to get them used to that

158 then after they have done the testing in the booths

159 you take them into the round table room then
160 and you write their results up on the board
161 probably do two or three weeks of that
162 or perhaps a bit of a mixture of project work
163 and a mixture of standards
164 to see how the variations have gone.

165 And then if that hasn't worked out well
166 you give them another shot at the training

167 KE

168 So this decision, is that based on one or more half hour sessions?

169 ... It is definitely based on more than one.

170 See some of them likes of Expert 3
171 she's been running a lot of half blinds.

172 Normally we probably run two or three weeks
173 it really depends on how good they are,
174 if they're not any good well you whip them back in
175 and work them back in the half blinds again
176 but whilst they're not two or three weeks,
177 it's just until they feel comfortable working in the booths

178 to get their confidence up

179 and assure them that yes when they look at their results on the board

180 they are still within, still even with the rest of them.

181

182 KE

183 So you are looking at six to nine sessions

184 ... Yes, you're really evaluating how good they are

185 and it's just a chance,

186 we usually have a look at them

187 and then get used to the actual project work,

188 it would be about two or three weeks

189 varies depending on the product,

190 a couple of years ago we did a study,

191 because the literature doesn't give you much guidance

192 on how to check panels variance,

193 it was really a measure that we decided on

194 that we could effectively use that was efficient

195 and it was something that we could do easily so we did a study,

196 we got a summer student to collate all the results

197 from every product that we had

198 and work out the average standard deviation

199 because there was no point in going for something

200 that was too low or wasn't even reasonable

201 and because we use different scales for different products

202 we had to work out standard deviations
203 based on a line scale
204 and a category scale wouldn't work,
205 so she found that on average
206 that the really average attributes like sweetness
207 or something that they do every day,
208 it could even go below one okay,
209 we found an average standard deviation of .5
210 for instance for sweetness in milk powder which,
211 I mean that is almost perfect given a group of people.
212 But some of them were over one on the ten point scale
213 so you know you're looking at ten percent fuzziness in the data
214 which is quite reasonable really,
215 so it depends on the attributes,
216 if they are not very common attributes
217 then they are very difficult going
218 and standard deviation might be something like 1.5
219 eventually we should have standard deviations for every attribute
220 some of the new panellist that are trained at the moment,
221 they are coming up with words that we have never used before
222 so we have got no history in which to gauge standard deviation
223 but depending on the scale KE,

24 like I said if it get any where near two on a ten point scale
25 it is a problem
26 no matter how much history you have got to base it on
27 if the attribute is always causing a variation
28 I would get rid of it,
29 because it is obvious not enough people are using it
30 and can relate to it in order to get that variation down

31 KE

32 What system do you use in the panels then
33 for the panellists to record and report the scores
34 about the product in use?

35 ... what happens in the training process
36 they start out with a blank sheet,
37 they move on to a sheet that's got a lot of words on it
38 and those words are eventually cut down
39 to what we call our final profile.
40 Then it's on those words that they do their scoring,
41 be it on a line scale,
42 which is chocolate at the moment,
43 be it on a nine point scale which is AMF,
44 be it on a ten point scale which is most of our products,

15 category scale.

16 So your panellists are either using numbers

17 or they are using a line scale with which to do their scoring

18 and it's from that that we then convert that into a mean score

19 with a standard deviation

20 so at the end of the day the panellists don't get instant feedback

21 as to what the standard deviation is on that day

22 but that will change once we've got our computers in,

23 that's the whole idea, instant feedback.

24 If you can say to a panel that day that they're way out of kilter,

25 I mean two weeks later it's no good hearing that news

26 they can't do anything about it then,

27 which is what's been happening

28 KE

29 Oh, I see. Why do the group leaders, why do some use a line scale,

30 some use a nine point and some use a ten point scale?

31 ... some people have an aversion to numbers.

32 Any scale that you use, there will be end-of-scale avoidance,

33 so your idea is to try and get a scale that's wide enough

34 to cover all possible options but also account for end of scale avoidance.

35 So if you've got a five point scale

36 you're only effectively going to get three numbers

67 used because everyone will avoid one and avoid five.

68 So that's one of the aims of our ten point scale,

69 that has enough, gives you enough discrimination and,

70 for them to avoid zero and ten

71 but you're still getting enough discrimination

72 on the rest of your points that are left.

73 Now Expert 3 uses a nine point scale,

74 I'm not quite sure... her and I have had long discussions

75 as to her using a nine point scale,

76 so you should ask her about the theory behind doing that.

77 A line scale is meant to get rid of this apparently

78 because with the category scale like a number scale

79 the assumption is made that the difference between one and two,

80 and the difference between two and three is the same.

81 It's not actually true

82 what we've used is what we call a 150cm line,

83 and we have intervals at 10mm and 140 or whatever it is.

84 And then, so you work out in terms of 150cm

85 but, yeah, you convert that basically into a percentage

86 when you start interchanging the scales with the product

87 we're going to run into problems with the standard deviation

88 On any sheet that you look at,
89 there's room for comment
90 and these qualitative people will always write something

91 we will take into account those comments
92 because sometimes they think words rather than score
93 and they'll say look it tasted really something or other
94 and you think why didn't you put the score up there.
95 So we'll often take into account,
96 now we don't actually manipulate the data,
97 because you get fired from Establishment 1 for doing that
98 but sometimes if, when we're interpreting the data,
99 if enough people have made this qualitative comment
00 then we'll actually use this information
01 more than the actual numbers

02 You can't do that with the selection process
03 the way we've got it designed at the moment
04 because we're not looking for numbers ability then

05 with copy sets we might find that a certain F ratio
06 or significant score is what we use as well

07 If they went a whole week

08 with an acceptable level of standard deviation
09 and I'd given them a wide enough range of samples
10 to cover most of the attributes
11 I'd be happy with them then
12 and if 2 weeks prior to that they were all over the show
13 then I might even extend the half-blind period.

14 3 full sessions of really good standard deviation
15 on a wide range of products, I'd be happy

16 I'd probably throw in a control every time,
17 something that's just standard
18 so that you're actually testing their consistency
19 cos you need to find out...

20 if you threw any duplicates in
21 then they should actually score it the same.

22 If they didn't score it the same,
23 then you'd be concerned

24 so you'd always throw in one reference every time

25 and I'd throw in, for the standard attributes

26 like sweetness, creaminess or whatever,

27 I always have something that is representing that every time.

28 And then I'd perhaps throw in one or two odd ones

29 like a really high sour one

30 or a really high oxidised one or something like that.

31 It will depend on what you've got on hand

32 but that's the procedure that I'd take.

33 So I'm really looking for...

34 I'd say 50% or more are standard type samples

35 you know routine stuff

36 cos that's mainly what they would be doing

37 and the rest of it just one off type stuff

38 of any other type of stuff they're likely to encounter...

39 over the week

40 if you have three really bland samples,

41 you know really high oxidised ones,

42 you are going to get a higher score

43 than if you had put 3 or 4 oxidised ones together,

44 so you have to take into account contrast bias as well

45 normally I would try for something, say a reference

46 and then sort of build it up in intensity

47 so you've got a little bit of oxidised, medium oxidised and high oxidised

48 or something, and they're all randomised anyway

49 so they're not all going to get them in the same order

50 if I'd gone 2 weeks and couldn't see any improvement

51 I'd say scrap this, let's go back

52 I'd say anywhere between 6 and 10 sessions
53 that's when I'd make the decision.

54 Cos, yeah if they're not capable at that stage
55 then they weren't ready to go in in the first place.

56 So your gut feel
57 or your standard deviation calculation was wrong

58 Go back to round table training

59 I'd take them back to standards, back to all the basics,
60 what you'd used as training for sweetness,
61 what you'd used for each of the attributes
62 to get that scale delineated again
63 because at the end of the day, we should have references,
64 say on a 10 point scale,
65 we should have a sweetness score of 0,2,4,6,8,10.

66 And I'd go back and sort out which one was a problem area,
67 which ones they're having the most problems with,
68 so I'd probably spend say 2 weeks on references again,
69 and then try throw some more project work at them again

70 We have changed group leaders because of that too

71 Her confidence in taking panels has gone down
72 so I actually sit in on sessions now and try and help
73 and this buddying system that I've been trying to get going
74 with the other ones as well to ensure that if someone is sick,
75 that the other person has much more confidence going in there
76 knowing what has gone on before

77 **Expert 2: excerpts by meaningful phrase**

1
2 we get people who respond to that ad
3 and we get them in for what's called screening sessions
4 which normally occur over a week.
5 They come to 5 sessions
6 where they will be asked to do a variety of tasks
7 normally includes identification of basic tastes
8 and may include doing what we call threshold testing
9 where they have to taste a series of solutions
10 until they can taste and identify the flavour,
11 but that isn't always done.
12 Secondly it will include normally triangle tests
13 which is where they are given three samples,
14 two are the same, one's different
15 and they've got to identify the odd sample.
16 In addition we normally get them to do some descriptive work
17 where they have little vials
18 which have got different kinds of food in them
19 and they've got to describe the smell or something like that
20 and we also do a couple of other things which are mostly for fun
21 and also to fill in time so they won't get too fatigued
22 We also try to look at the person's personality,

23 try and weed out any obvious undesirables
24 and so in the most recent lot of screening
25 we actually had a small interview where each person
26 ... just trying to find out what their motivation was...

27 Then at some stage we set a level
28 above which we will accept people and below which we won't...
29 We normally allow for more people than we actually want
30 because we get, we always get a percentage who are unreliable,
31 not interested... a whole bunch of things...
32 and then from that the training starts.

33 In my mind there are two different things that we have to do
34 because sometimes we screening for a total new product
35 that we haven't done before
36 and in other cases, which is what I'm doing at the moment,
37 we have an existing panel which we need to add people to
38 so I have to try and train those people
39 to fit in with the existing one
40 and that's a little bit different
41 to training for something completely fresh and things...

42 If it's a completely fresh product which we haven't done before,
43 don't have an existing panel for,
44 we normally would spend a few sessions, maybe 3 or 4

45 just getting into word generation
46 so you give them different examples of the product
47 and get them to just try and describe it
48 and try and steer them away from subjective words
49 like bland or nice or yuck or whatever
50 to actually trying to say well you think it's nice
51 but what about it that makes it nice
52 or what about it makes this thing disgusting.
53 Things like that. At the same time in the introductory sessions
54 we go over things like how to taste,
55 how to spit out nicely in public, yeah,
56 and a little bit about what the section is about
57 and what the institute is about.
58 Need to try and give them a little bit of knowledge...

59 Trying to group the word together into broad categories
60 and then ultimately agreeing on the one or two words
61 to describe that type of flavour.

62 Following on from that we start trying to develop reference samples
63 for each of those areas
64 so like for sweetness you have to find something that's not sweet
65 right through to the limit of what we could expect that sort of thing,
66 and hopefully some points in between...

67 At some point along there we'd start introducing some kind of scale
68 depending on what we wanted to do with them.

69 At present we're using what's called a category scale,

70 don't know if you're familiar with that term,

71 like scoring something from 0 to 10 is sort of like category scale

72 so we might start them off by trying to use words

73 like slight, weak, moderate that sort of thing

74 and then transfer them onto numbers when they've had a bit of practice

75 at trying to put it into categories.

76 Another kind of scale which we use for some panels is a line scale,

77 150mm long with an anchor point at each end.

78 So with that we'd start them off with that rather than using words...

79 we're starting to try and find things which are reproducible

80 and are reasonably consistent for the panel...

81 Along the way, you should be trying to give them feedback

82 on their individual performance...

83 it's something that we're going to add onto,

84 we do basically the same things

85 but we have to try and guide them a little bit

86 so that we don't end up with 2 different lists of words...

87 So we'd also be getting the existing panellists to join that group very early.

88 Basically as soon as they've got over their first couple of weeks,

89 they've just got their words and are trying to put them into groups

90 so that we can make sure that everyone understands the standards
91 and if necessary we will use some of them....

92 KE

93 Mmm. It must be a lot harder to get reinforcements onto a panel
94 than to build a whole new panel... to get consistency.

95 A lot harder, because you tend to end up with factions.

96 KE

97 Yeah, the dynamics are disrupted.

98 At what point do you declare a panel void and start again for a product?

99 That almost happened for the chocolate panel because, for two reasons.

00 One, they got very low in numbers fairly sudden

01 and secondly that we started to get a really big variation in their data...

02 I didn't train the panel originally so it has been quite hard taking them

03 over because it seems to me that perhaps they didn't all agree

04 on some of the definitions or their understanding of them was not clear...

05 normally it's a matter of going back

06 and trying to fix that before going on and doing somebody's project...

07 KE

08 How many is not enough for a panel?

109 Less than 8...

110 we start off with about 20

111 and they sort of whittle down within the first week or so

112 to about 16 or so. We tend to lose a few.

113 Around 20 to 15 is a good number.

114 We can then get them... if they're going away on holiday

115 we don't drop too low.

116 If you only have about 10 and one gets sick, you can really have a problem.

117 Yeah, around 15 is a good number. 20 is hard to control...

118 In theory, we declare it trained

119 when they can reproduce or at least evaluate the same samples,

120 when all the standard deviations are within the limits that we've set

121 and sort of remain there quite happily.

122 A third factor that comes in quite a lot

123 is the pressure to have them ready...

124 The group leader. I mean I would discuss it with Expert 1

125 but I'm sure what she would say is, if you don't think they're ready, don't do it.

126 We might decide that we'll do a round table discussion

127 but we won't do any statistics or anything if there's real pressure on.

128 That normally is the group leader's decision...

129 went back through whole heaps of data

130 and worked out what the average standard deviation was

131 for the different attributes and then we tried to set limits

132 that we thought were reasonable and achievable

133 but not so easy to reach that they were nonsense.
134 Most products or most attributes it's plus or minus one on the scale
135 but some of them go up to 1.5,
136 for attributes that we know the panel have more trouble with.
137 So yeah, you've only got a 10 point scale
138 so when the standard deviation starts to get too high
139 it sort of makes a nonsense of it, of the data.
140 you want to show that the panel is able to score
141 using the same part of the scale and can do it consistently...
142 a really simple product I would say 8 weeks is really the shortest,
143 that's with them coming to 3 sessions a week.
144 And the longest... oh 6 months maybe...
145 things like baked products.
146 I anticipate that chocolate is going to take quite a few months.
147 3 is sort of about the average.
148 Things like butter and whole milk powder normally run for about 3 months.
149 You tend to take longer with a product
150 where the flavours are more sort of tied up together,
151 and it's really hard to break them apart and treat them individually.
152 Things like yoghurt. The flavour is reasonably simple...
153 some of the things like chocolate and baked products
154 it might be several weeks before we get those samples
155 and we've got to go back to what they were talking about 2 weeks ago.
156 The chances are they will have changed their minds anyway,

157 it's a much more difficult process...

158 Something like chocolate, we've got what we've got

159 and if we want to try something out it might be, you know,

160 by the time it has been manufactured it takes 3 weeks before people taste it

161 to for it to set up properly

162 and it might be 3 weeks before there's time available to make that sample.

163 So any product that's liquid we can prepare here,

164 and that really helps the training process

165 because you can listen to what the panel's stating one day

166 and say tomorrow we'll try x, y and z, and do it...

167 time that it comes out is at our performance review,

168 like if I took a whole year say to train a chocolate panel

169 then that would come out in my performance review.

170 Okay while I might have been doing everything that I could think of

171 it probably would be that I should have been seeking more help

172 or getting more support or... admitting defeat.

173 I don't know, because to achieve at least half of the objectives that we have

174 your panels need to be performing

175 and if they're not then those objectives will suffer

176 and that comes back in that form...

177 normally 3 sessions a week

178 and they last anywhere from... they can last up to an hour...

179 a big group they take longer

180 and like, for a panel that is trained like the butter panel

181 they still have about one training session a week or a fortnight,
182 depending on their schedule.

183 Some of them are quite quick, 15 or 20 minutes
184 because there isn't that much... it's more for them
185 so they can see how they score compared to everyone else.

186 I would say that the average is about 40 minutes...

187 The main reason as far as I can see that we have them on regular times
188 and days and things is so that they get into a routine...

189 if you only did 2 sessions a week

190 you didn't actually achieve that much

191 because you might start off on say Tuesday working on caramel flavour
192 and it might take 3 sessions or more to work through that.

193 If you only have 2 then a big long gap,

194 you will have to repeat what you did on the last day.

195 4 we decided was asking a bit too much of people.

196 we didn't think there was enough dedication in the community to get that,
197 so yeah, 3 was a trial thing but it works really well...

198 We pay them \$10 per session

199 regardless of the length of the session so if we do 10 minutes or an hour,
200 they get \$10. They get paid by cheque fortnightly.

201 Well technically they're not paid, it's reimbursement
202 so they don't have to pay tax on it.

203 Legally they're volunteers.

204 Also with some products we bribe them with food.

205 Like the butter panel might get chocolate afterward
206 because when they first started they all moaned terribly
207 and were really offended that they would have to eat butter
208 and other people got to eat cake and chocolate...
209 They get paid right from the time they come to screening.
210 They get paid less for coming to screening.
211 Normally if they came to a 5 day screening they get \$30.
212 And they get paid right through their training, that sort of thing.
213 We're also starting to introduce some sort of bonus scheme,
214 not on a regular basis but just every few months
215 for those people that have attended more than say 90%
216 then they'll get a small bonus
217 like a book voucher or a petrol voucher or something...
218 for the initial setup of a panel, you can have multiple clients
219 and Establishment 1 in fact will be paying for that,
220 that'll be as part of the working expenses of the unit
221 but then if we go training for a specific project
222 then that's project work...
223 I'm responsible for 2 panels,
224 one of which is trained and the other is in training...
225 One of the things that we do in training
226 is put all the scores up on the board
227 and everyone can see what everyone else has put.
228 Someone who is a non-performer tends to show up then

229 and normally I would talk to them and say,
230 look KE you put 2 for this item and everyone else has put 10,
231 perhaps you'd like to go back and try again
232 and see if you can understand what they're talking about
233 or if there's some other flavour or some other way you'd describe it.
234 It's very difficult though because often when you do that,
235 people go back and say, yes I can get it now
236 and you really don't know whether they can or not...
237 we've done various projects for people and the results have been confounded
238 by the variability of the panel
239 and it looks like there's something there
240 but there's so much variability in the data
241 that you can't draw that conclusion confidently,
242 and so you've got a client saying, look,
243 we're paying you x thousand dollars to do this
244 and you can't tell me because your panel is so variable
245 that you're not sure. And that reflects badly on me and on the section
246 and on sensory as a whole
247 and so to improve that variation within the panel has to be dealt with...

248 KE

249 When you've got a person that's performing well,
250 you trained them up and then performance starts to decline, what causes it?

251 Probably depending on the individual.
252 Some things are that they get bored, on a tasting,
253 they sit down, write down a few numbers, take the money and go home again
254 sort of attitude. Maybe not enough on-going training
255 like if I put them in the booths week after week
256 and don't give them any feedback.
257 People may start going off on their own tangent.
258 Stress, pregnancy... got one of those at the moment...
259 Every panellist has their bad days or bad day or two...
260 panellists have actually commented that...
261 see often a lot of the samples are really similar...
262 and after maybe the second week of it they sort of start to think,
263 gosh there must be something wrong with me, these all taste the same,
264 you know, I must be missing something,
265 and they start getting really paranoid...

266 KE

267 What do you do with people who are good performers
268 but through fatigue or adaptation they're just losing their sensitivity
269 to a certain product?

270 I guess they're probably quite difficult to identify
271 making sure that they're remembering what those reference samples are
272 and they're scoring things according to those samples.

273 Panellists have a real tendency to want to score samples
274 relative to one another on the day
275 and so therefore you can never get day to day consistency.
276 So yeah, going back over those reference samples regularly,
277 I think can help stop people becoming desensitised
278 but you also get the opposite effect that they can get over sensitised
279 and tell you that they can taste a hundred and one different flavours
280 in the sample and maybe they are or aren't but no-one else can taste them
281 or you know... yeah, I think reference samples are about one of the best things
282 we can do...
283 we still limit it to the three or maybe four sessions per week
284 and for most products we limit it to about four samples per session...
285 it starts when we have enough requests programmed
286 waiting for a certain product type to justify setting up a panel
287 and the decision is made to do that
288 and someone has to be responsible for it...
289 you're responsible for producing that data or having the data produced
290 and reporting it and being accountable for it,
291 in that you can understand, explain and justify what was said about it...

292 KE

293 So is the screening generic or is it related to the product?

294 It's a mixture. Like a generic base but with things that can be added on

095 or taken off depending on what it is we want to achieve...
096 in some instances we just know that say there's going to be heaps of cheese work
097 and we're going to set up a panel
098 and the project information comes afterward
099 or during the training process
100 and other cases we know before we start
101 like Technician 2 and Expert 4 are the classic examples.
102 They know exactly what it is that the panel is required to do...
103 you have to sort out times of the day when you're going to run your screening,
104 how many people you want to attend,
105 how long it's going to take them,
106 the equipment that's going to be required
107 and the samples and everything required.
108 Any publicity information you want to give them.
109 Who's going to do what tasks on the day...
110 During screening we have 2 approaches.
111 One is that you can plan it all out and say right this is what we're doing
112 on days 1 through to 5 end of story.
113 Or if you're about to do something that you don't have much experience in,
114 what we might do is say plan the first two days
115 and go through the results
116 and say well the screening is far too hard.
117 No-one can pick the differences between samples, it's too difficult.
118 By making it too difficult you're not going to achieve anything,

319 because you're not going to select anyone.

320 So we'd normally want to try and find a level which would spread people out

321 and show their different abilities to discriminate Ok.

322 So you've got to try and find that level fairly quickly,

323 because we only have 5 days to do it.

324 We have a lot of base knowledge with which to do that,

325 so it's setting up the actual screening...

326 KE

327 You mentioned earlier on that you interviewed you potential panellists.

328 Who does that?

329 The panel leader does.

330 I had a look at their data

331 and sort of pencil in those people that I thought had potential

332 and so trying to take particular note of them

333 and to see what they'd be like...

334 Interviewing is an aspect of selection....

335 group co-ordinator.

336 Actually standing up there and trying to see the wood for the trees.

337 The other one is more of a physical task I guess

338 like organising the samples,

339 co-ordinating them,

340 trying to select samples which might help the group in what we're trying to say.

341 At the same time reading the literature or past reports or whatever

342 and trying to understand that from a food science point of view

343 but also trying to be a facilitator, you know,

344 without trying to put the answers into their head...

345 One part of it is the researching side,

346 trying to understand any reactions that might be going on in the product,

347 reacting compounds and that sort of thing.

348 The other one is maybe a little bit more on the technical side

349 and may actually be delegated to some one else,

350 in that it's the physical handling of samples,

351 checking out sample availability,

352 preparing sheets for training

353 because often during training we might go through half a dozen sheets or more...

354 There's the analysis and interpretation of the data

355 that's being generated during training..

356 KE

357 Right. They're all interrelated

358 because otherwise it wouldn't be one person's job.

359 Could we say that analyst is a task?

360 Yep

361 General communication, like communicating with Expert 1,

362 communicating with clients,

363 communicating to the other staff here as required...

64 **EXCERPTS - SESSION 2**

65 giving the panel confidence
66 and showing them that they are doing well
67 and that their scores are meeting whatever criteria we've set,
68 whether it's standard deviation or range or whatever we've chosen...
69 giving them feedback sort of like that day or the next day
70 on how well they did and if we repeat that part,
71 show them what their scores were over certain days or whatever,
72 just so they feel good about themselves...
73 be confident in myself that each of the panellists do know what the attributes are
74 and that they can score them reasonably well
75 and that should have been achieved by the use of standard samples
76 and things...
77 start bringing in project samples
78 so if cake is going to be something unfamiliar
79 which they haven't come across before, which quite often happens,
80 or it's like what they've had before but because it's not a standard sample
81 which we've manipulated, it's not a naturally occurring flavour or whatever.
82 It might be in the same flavour group but not exactly what they're used to.
83 They have to learn to recognise that and know where to put it,
84 rather than sticking everything in the 'other' category
85 because it's not exactly as they've learned it...
86 don't normally tell them too much about the aim of the work or anything
87 like that because it can result in quite a lot of bias if you do that too much.

88 It's difficult but it's good to try and find some sort of balance
89 so they know a little bit about what they're doing
90 but not so much that they then skew the data for you...

91 KE

92 That's good. In relation to you actually confirming
93 whether they understand the attributes and can scale them or not,
94 what actual range of tests do you put them through to do this?

95 Traditionally, it's just a round table training session like in here
96 and yeah, giving them reference samples
97 and making sure people can recognise them...

98 like in training they might be doing say a flavour which we call 'cow-ey'
99 and they might have had it at three different levels, OK

00 and they get those often in a similar order

01 and if you suddenly throw one in amongst lots of other samples

02 they still should be able to recognise that...

03 The other thing to do is to put them in the booths

04 and make them do it by themselves

05 because a panel that does all their work in here for 3 months

06 and all of a sudden they have to do it without any visual feedback,

07 they often just fall apart for a couple of days...

08 in every panel, there's always at least one attribute

09 which I'm a bit uncomfortable with cos I think the panel flounders a bit.

10 So I might give them more leniency on that
11 at the end of the training, I'd expect only to have to do it a few times
12 because their training should have been good enough,
13 but I have confidence in that and it's just a final check.
14 It isn't something that we do that formally,
15 just the last few sessions we take particular notice
16 and maybe give them some of their data back and that sort of thing.
17 Point out to the people if they're consistently scoring lower than everyone else...
18 at the end of it, they should be able to focus on every attribute in the sample,
19 saying is it there, isn't it, how much...
20 When I looked at the data, I'd look at every attribute,
21 but the sample that I gave them might have focussed on one or two...
22 depends on the panel, and the attribute and the product.
23 Normally we like to have the standard deviation round plus or minus 1,
24 that's on a 10 point scale.
25 But like if we have an attribute which we know is the weaker points of the panel
26 then we can allow but 1.5 is pretty much the end of the line.
27 Getting above that makes your data pretty fuzzy...

28 KE

29 Could we just go back to the testing phase,
30 the bit at the end of the training cycle.
31 What happens if you have the gut feel that this is the right time for them to finish,
32 or there's a bit of commercial pressure from a project sitting there,

33 and the performance simply isn't there,
34 like the standard deviations are beyond 1.5.

35 Sometimes the work goes ahead anyway
36 but normally we tell the person,
37 look, we can do it but the panel, you know, isn't up to scratch
38 but if that's what you really want to pay for then, you know...

39 but normally, no...

40 if they had standard deviations of 3 or 4,
41 then you would have to look at are the reference samples inappropriate
42 or maybe you've got some panellists that just aren't up to the task.

43 I mean, when it gets to a situation like that,
44 you really have to start looking hard at individuals.

45 Is the standard deviation high because everyone is putting 2 or 3
46 and one person's putting 8.

47 Or is it because the panel is completely spread over a wider range...

48 If they are consistently causing the performance of the panel to be low,
49 then they need to be told that,

50 and if it's something that they can't help

51 then they either have to learn to adjust

52 or else they need to be off the panel...

53 KE

54 So do you keep individual records of the panellists

155 to track their performance?

156 No, we don't.

157 Sometimes, like for a specific project, I might do an analysis

158 and look at individuals

159 but it's not something that we do standardly

160 although I hope that when compusense is installed,

161 we will be able to do that a lot more easily...

162 The first step is to do more training on it

163 and to try and find out from the panellists

164 if they think that they really understand

165 what that attribute is because particularly here I think,

166 there's a real tendency at the start

167 for the panellists to want to please you

168 and so they say, you give them a sample,

169 and you say we thought the sample might have whatever characteristic

170 and they go yes, yes, yes

171 and if you think yay, I've made a breakthrough

172 and if they haven't been completely honest with you,

173 then you might think that they understand that standard

174 and everything's home and hosed

175 it has caused problems with the chocolate panel

176 because they get an unknown sample like in project work,

177 they suddenly don't know what it is

.78 and they don't know how to score it
.79 because they've learned the process the wrong way round.
80 You have to ask the question,
81 is it because somehow you've tried to impose
82 a word on them that they're not comfortable with.
83 Is it that the references aren't appropriate.
84 Is the attribute even really relevant to the product?
85 you have to go back and look at what you used as training standards,
86 whether with hindsight something else could have been more appropriate,
87 whether in the experimental product
88 it's maybe a sub-group of that flavour
89 that's coming out or whether it has gone beyond that flavour
90 into something new
91 If it's one or two individuals
92 then you have to look at whether they have at the same rate
93 as the other people or as well or have the same understanding,
94 are they really understanding the scale
95 or are they even understanding what's being asked of them.
96 And if not then you have to make a choice
97 as to trying to fix it or to take them off the panel
98 how many people have I got rid of?
99 only about 2 off one panel.
00 Expert 3's got rid of... sounds terrible... taken off the panel
01 I think 1 person off the baked products panel

502 if they have some sort of chemical blind spot
503 in which they're just not capable of perceiving the chemical
504 that's causing that flavour
505 Another quite likely cause is that, for example in chocolate,
506 we measure burnt flavour
507 and there's always been one or two people on the panel
508 that the first time we give them a burnt sample they say it's nutty, OK.
509 Don't ask me why but it has happened.
510 And if I as a panel leader
511 or that person fails to recognise that connection,
512 then that person will carry on scoring that incorrectly
513 a lot of it's panel leader error or interpretation,
514 and yeah, I would say that one of the most likely causes
515 is through the training
516 The panel leader has to be really tuned in to that kind of thing
517 in that if you have a panellist who is consistently not perceiving something,
518 instead of just writing him off as useless,
519 trying to see, is there something else that's going on
520 I can't think of any formal checks that we have,
521 on how the panel leader has done other than how it's reflected
522 in the data and group.
523 And that doesn't mean to say that if the data isn't that good
524 that the panel leader has necessarily done a bad job

525 that person has a responsibility then to ask for more help
526 or get more support from within the section
527 most panels we have tried to have like the panel leader
528 and someone else in the section
529 who knows a reasonable about what's going on with that panel

530 a panel is never perfect
531 and I don't think I'd say that they're never trained
532 but I'd say there's always room for improvement
533 so the work I'm doing at the moment,
534 I'm making sure that their data is being collated every day
535 after the panel
536 and I'm putting the mean and standard deviations on a graph
537 out for the next panel
538 so if I need to I can repeat that day's tasting
539 or I can talk to individuals who may be having a problem.
540 or if the whole panel has gone off the edge then we can stop it,
541 have a few day's training,
542 talk about the problems and go back and start again.
543 When it happens during project,
544 the problem arises if you're not on top of it
545 and that's something that we haven't done that well in the past
546 but we're starting to be a lot better at it
547 because if you let the data mount up

548 for three weeks while you're doing project work
549 and then find that half the days are absolute garbage,
550 I mean OK you could go back and start again
551 but a lot more extra work and a lot more expense so it's, yeah,
552 you're much more in tune with what's going on
553 if you're looking at it daily
554 compusense measures it in pixels
555 and it's limited to 61 pixels,
556 whereas before we had a 150 mm line,
557 so we have reduced the sensitivity of the scale by more than half
558 and that's quite a concern to me.
559 it might make my standard deviations look better
560 but I also think that for the panellist that it's much harder for them
561 to get the sensitivity that they feel they've got on the scale
562 Within 2 or 3 minutes after the panels finished
563 you can get mean and standard deviation, range
564 and you can get them to wait in here and show them
565 if you wanted to make it a very clear cut decision
566 then we need to have set standard deviations and perhaps ranges,
567 some kind of defined tool to measure them by
568 we don't have it formally defined how many days in a row
569 they should be able to do this
570 or like with duplicate samples,
571 how close they should be, like within a day or over a day

572 so yeah it probably needs a little bit more definition
573 if you wanted to be very strict about it.
574 But also it's difficult because there's also a subjective side to it
575 like if it's just 1 person that's messed up, does the whole panel suffer
576 because of that?
577 and if they don't do it again for a week
578 and then it's another person, do you keep holding them back
579 because of that
580 so it's almost like you need a band with a critical level on either side
581 when they finish training and begin project work,
582 tend not to tell them because
583 it creates this great excitement
584 and they're all expecting... different samples
585 that often are very standard and very routine
586 and so there's a psychological thing

87 **END OF EXCERPTS - EXPERT 2**

Expert 3: excerpts by meaningful phrase

1
2 we advertise and as a result of that we get 60 to 80 people
3 and you set them tasks.

4 Well, when I'm running the screening I set them tasks
5 that are related to the product that they'll be evaluating.

6 We go through, first of all, the basic tastes
7 and go through some difference testing
8 and then some difference testing with the product itself
9 using project variables to make different samples
10 and that's supposed to get harder as the work progresses
11 and they get more used to it

12 but, yeah, it's quite difficult

13 I'd do a bit of personality work as well,
14 just group dynamics

15 And then training starts

16 and you just basically do an introduction
17 to get them all familiar with each other,
18 you know where Susan comes from and all that
19 so they can get down to what they're doing then
20 just vocabulary development,
21 that's probably the first stage

22 where they generate lots of words

23 to describe lots and lots of different samples

24 and you try to get samples as varied as possible at that stage,
25 so that they can really get any sort of attributes that they can think of.
26 Then later on, if you're just doing flavour or just doing texture
27 you start to concentrate on those.
28 We probably use the same samples to begin with over again.
29 Then introduce a preliminary, basic scale
30 like absent, weak, moderate, high
31 and try and pick out individual attributes and work on those
32 So if they all agree that their milky is say similar to someone else's creamy
33 or they all agree that it is present in that attribute
34 then I'll put to the panel that I think that everyone
35 is calling the same attribute different words
36 so what can we call it that everyone will recognise.
37 And you go through that and that's a really circular thing
38 then you start to work on scaling each attribute
39 with that basic, preliminary scale
40 and then trying to get it to be a bit larger
41 but at that stage a few decisions have to be made by me
42 regard what type of scale to be used
43 then setting up standards along the scale I suppose is the next stage.
44 So the aim of it is to have them all understand each attribute,
45 understand the scale
46 and be able to recognise specific samples along that scale
47 for that particular attribute

48 put them into the booths
49 and do a few tests to see how consistent they are
50 They've just shown interest from filling out a questionnaire that we have,
51 they go on to our database and then, yeah,
52 if they're available at the time of our screening
53 and they're going to be available for the next however long
54 then they're screened
55 then based on their performance in the screening test
56 they're put on to the panel.
57 That's if they're 75% correct... On the tests that they've performed
58 for the week's screening.
59 In the two panels that I've done they've had two weeks of screening
60 like they might get 8 samples and they have to taste all 8
61 and put them into order of increasing sweetness.
62 That's one sort of test they can do
63 or threshold test where they've got a sequence of 8
64 and they have to say when they can taste a difference from the reference
65 just a detection threshold really,
66 that there is a difference
67 and then a recognition when they recognise what the actual taste is,
68 that's one other test.
69 Difference testing is used which is,
70 well there's several different ones
71 but the common one that we've used is triangle tests

72 which is with 3 samples and you just circle the sample
73 that's different to the other two.

74 The description tests we do with a sample
75 and they have to describe the aroma

76 The last 2 times they did interview them.

77 I don't know what they asked them or anything like that.

78 I haven't interviewed any panellists

79 I think I'd run a training session like with bizarre samples,
80 different sample you know and just put the group together

81 just to see who really takes over or who didn't

82 I think it's quite important to have people that are really neutral,
83 who are willing to learn

84 KE

85 Mmm. So how long would you have them
86 and for how long in the selection process?

87 Every day for 5 days...

88 For about an hour.

89 Yeah, but because there's like 60 to 80 people
90 it would be a whole morning.

91 You'd have groups of people coming in and going all the time

92 I'd say 15 to 20 is what I'd choose

93 cos there's always quite a rapid drop out rate in that first 3 months,

94 people that don't like it or don't want to make the commitment,
95 that sort of thing so yeah,
96 and then I'd expect that to drop down to about 12 to 14
97 To run a panel you have to have 8 panellists...
98 You can't work it out parametrically if you've got less than 8
99 Well they'd come in for the preliminary first week
00 and basically the first day is... well probably the first week,
01 we really introduce them to what they have to do
02 and we really push the things like don't wear perfume,
03 don't clean your teeth before you come, you know,
04 you're just reinforcing the principles like
05 just do your own thoughts and all those things.
06 And also start looking at any varied samples.
07 And on the first day they introduce themselves to each other
08 and give a background
09 and I just then start them off writing down as many words as come to mind
10 about each sample that's in front of them
11 I'll tell them what each sample was
12 and talk about it a bit and then probably the next week,
13 start looking at the words that they've written up there
14 and ask them if they can see any words which could be similar.
15 At that stage you just start varying all the samples that you can think of
16 and if there's one word which keeps coming out a lot,
17 you stick that down on the form, so they've got that to use...

118 It's at that stage that I get caught up in trying to produce samples
119 that are going to be accurate enough...
120 Accurate enough to get everyone to agree to a particular attribute...
121 the more complex the product the more horrendously difficult it is
122 to get accurate attributes
123 we have got just a general thing on sensory evaluation
124 and it has got things like the tongue, you know,
125 where the stronger areas on your tongue are,
126 not to brush your teeth and that sort of thing.
127 I can't remember everything that's on that handout.
128 It's just really general stuff but on the back of it is a list of things...
129 I'll introduce a scale
130 like probably the first scale that I would use
131 would be absent, weak, moderate and strong
132 so they've only got the A, W, M or S for each attribute
133 our preliminary evaluation form?...
134 it has got the list of the attributes,
135 it has got the scale
136 and they just have to put the capital letters that they think are in there
137 I try and get people to interact as to how they're tasting things...
138 It doesn't actually last that long with that sort of scale.
139 We just try and get them to go through with each of the attributes,
140 then I normally introduce them to a category scale
141 in any of the work that I've done.

l42 The primary reason for that is that historically
l43 that's what Establishment 1 has used
l44 it's a 0 to 9 point scale and they can only use the points,
l45 it's not a line scale,
l46 and the distance between points is not set so the points are anchored,
l47 like 1 is the threshold level, 3 is weak, 5 is moderate, 7 is strong and 9 is intense
l48 but there's no actual linear, no mathematical distance between the points
l49 basically what I do is I go through different samples,
l50 getting them to practice on the scale
l51 and just discussion until they've got a good grip of what they're doing...
l52 I'm analyzing the results throughout the time
l53 and giving them feedback about how they're going.
l54 So basically that's it until they're competent at each attribute
l55 and we've got a system where, I don't know if Expert 1 showed it to you,
l56 where we've calculated the standard deviation about the mean for the panel.
l57 If they're too high, if there's too much deviation,
l58 they're not ready
l59 we redo that work where the standard deviation is too high

l60 KE

l61 Right, and what is the formula?
l62 or is there a set of data that you're comparing against?

l63 Just whatever seems realistic

164 KE

165 How long then does it take you to get a trained panel?

166 6 to 8 months I think.

167 Milk fat took 6 to 8 months.

168 I mean the literature says 2 to 3 months

169 Feedback of their results but also getting real work.

170 They love seeing their own progress.

171 They love seeing graphs of themselves and what the average was

172 and what the standard deviation was

173 and who was too high and who was too low

174 if they're consistent, regularly turning up,

175 we give them money and things like that

176 KE

177 Yeah. What causes performance to drop away other than lack of motivation?

178 Lack of variation in the samples

179 **EXCERPTS - SESSION 2**

180 Well I start off, I just have a vague idea of what sort of product

181 and the particular attributes

182 so I start off and I just get really different,

183 like if I can get samples from overseas that are really different
184 I'll get them to profile things
185 and just profile standard samples
186 that they'll probably be getting in project work
187 and then I'll probably, in their results
188 I'll just have a glance at their results
189 and see that there is no-one right out of the ballpark
190 but probably by that stage that's not likely
191 because they've been weeded out to a certain extent
192 I'll do an ANOVA between panellists,
193 and then I'll look at the mean and the standard deviation
194 and if the standard deviation is quite high,
195 then they still need training
196 because there's too much noise around the mean.
197 If there's one person who's outlying on a particular attribute
198 then I'll get them, you know, work with them on that attribute
199 and do a training that's specifically geared up for that attribute
00 so I can see how they perform.
01 I give them some stats analysis like principle component analysis
02 but I've found that lots of those multi-variate analysis methods are too generalised
03 and it's just looking at their consistency really.
04 If I do the same sample twice in a row
05 I see whether the same panellists have given them the same scores

206 Over 2 days but that's probably what I'd do and if that all checked out.
207 There's no formal sort of stages that they go through
208 but if I get a feel that I've got confidence in them,
209 basically they'll do
210 We'd probably do each attribute 2 times
211 low standard deviation and consistency in the attributes.
212 if I were to change from a 9 point to a 10 point scale
213 and the acceptable standard deviation was 1 in 9,
214 then I'd just take that percentage and apply it to the 10 point scale
215 so that would work out to be 1.2 or something
216 If I had the training and they're all over the place
217 and they don't understand the samples
218 then I'd say come back tomorrow and I'll see if I can get more...
219 like we have the training
220 and they might not go very well on one particular sample
221 and it's quite an important sample or something
222 and I'll think I'll get them back next week
223 and we'll still have this sample perhaps with a different angle,
224 so that's what I'd do first
225 and if they still can't get it I'd just take them back to basics again
226 and start them on a really wide variety of samples
227 and then try and tune it down again
228 then 6 weeks down the track after you've done 2 more blocks
229 you might go back to the block that you did 3 times ago

30 and see how they're still going with that,
31 that's more realistic
32 I don't tell the panel what to taste sort of thing.
33 If they're saying to me, we just don't understand it
34 then I sort of drop it more than anything else
35 drop the whole tack that I'm on at that time,
36 try and change tack and maybe try and go around the particular attribute
37 that I'm working on
38 I just basically kept training them and training them
39 and changing tack and repeating things
40 trying to get things that can taste and stuff,
41 to understand it
42 I'd put them in a panel just to see how they'd go
43 and I've probably trained them with the samples that they'll be evaluating anyway
44 and if they just way out of the ballpark on them
45 then I'd put it off probably.
46 But yeah, that has it consequences
47 but rather than get results that are quite fuzzy
48 If I had a panel that was consistently not performing
49 I'd go to Expert 1 and I'd explain to her.
50 What would Expert 1 do?
51 Expert 1 would probably ask me if I'd done a lot of things,
52 suggestions to try and get them back into line
53 and I'd probably go away and...

254 I'd to be able to make the decision
255 that they're worth it or they're not worth it
256 but I'd have to make that decision with Expert 1
257 based on my recommendations I suppose.
258 I try to get real samples
259 like I try to get things that have been rancid for 8 months,
260 and then something that's moderately oxidised
261 and put them in the booths
262 and see whether they still say that they're moderately oxidised.
263 It's sort of not doctored samples
264 because my interpretation of doctoring might be wrong
265 for that sort of thing so I try to make them as real as possible
266 Probably a week, I'd say a week of panels going into the booths,
267 coming out and training straight after that
268 and probably be a week with the samples
269 that they'll probably be using in the project
270 plus some of the old references that they'll be familiar with
271 They'd probably be duplicated
272 so that I could see if they were consistent within each person
273 and then I'd take them back into the round table room
274 and make sure that everyone recognised that the lactic acid,
275 well the different flavours that were in the lactic acid
276 were more or less standard

77 what I've tried to do is suss out the parameters of shortbread flavour,
78 that's what we've been doing.
79 I've worked with the flavour chemist,
80 and he's been trying to enhance specific attributes within the shortbread
81 and I think that they're just starting to understand specific attributes

82 KE

83 Right. How many times would you accept them doing...
84 if the standard deviations are all good across the group
85 for a certain attribute,
86 how many times do they have to produce that result
87 before you can say, OK, they understand what butterscotch flavour is?

88 Some of them, in excess of 3 times.

89 on one particular attribute it goes from 0 to 8,
90 I'll say to the people who put 8, why have you put it so high,
91 please explain to everyone else.

92 They'll explain it the way they thought it,
93 and then I'll say to the person who put 0,
94 now can you relate to anything that person said
95 and they'll say yes or no.

96 I'll say why have you got 0,
97 what was it about the sample that made you think that there was none of that
98 particular attribute.

299 If they can't agree
300 then I'll ask everyone else if they can understand what's going on
301 and if they can't understand it...
302 I mean quite often when it's 0 to 8
303 it means they just don't understand what the attribute is
304 I used to have a cooked butter and an uncooked butter flavour in shortbread
305 and uncooked butter was consistently all over the place.
306 They'd say they could taste uncooked butter
307 and when I got them to talk about it
308 they didn't know what they were talking about,
309 they were just talking through a hole in their head so I just took it off.
310 Probably what I would do is just take it off.
311 I said to them, if I can find a sample that typifies this,
312 then we'll put it back on
313 but until I have a sample that you guys can consistently find uncooked butter
314 flavour, then I'm not putting it on

315 KE

316 In your experience, how long do they tend to go on for,
317 from the time that you sort of get a gut feel that they're about ready,
318 how long does it go on from that point to you saying they are trained?

319 I take ages to train my panels.

320 Probably 6 weeks

21 I told them there's none there,
22 because everyone else has said,
23 they don't understand that
24 so they've got to understand each attribute and each scale
25 which I think they need to learn themselves

26 **END OF EXCERPTS - EXPERT 3**

Appendix E

Concept Comparison Chart

Combined concepts chart			
NO.	EXPERT 1	EXPERT 2	EXPERT 3
STAGES OF TRAINING			
1.	Selection	Selection	Selection
2.	Training	Training	Training
3.	Project work	Project work	Project work
4.	Decision point	Decision point	-
TRAINING ACTIVITIES			
5.	Threshold testing	-	-
6.	Triangle tests	-	-
7.	Product description	[Word generation]	[Word generation]

Combined concepts chart

NO.	EXPERT 1	EXPERT 2	EXPERT 3
8.	Words to numbers	Words to numbers	-
9.	Feedback	Feedback	Feedback
10.	Group dynamics	[Group co-ordination]	Group dynamics
11.	Training session	Training session	-
12.	Analysis of performance	Analysis of performance	Analysis of performance
13.	Half-blind tests	-	-
14.	Scale development	Scale development	Scale development
15.	Product evaluation	[Reference sample development]	[Reference sample development]
16.	-	Panel protocols	Panel protocols

Combined concepts chart

NO.	EXPERT 1	EXPERT 2	EXPERT 3
17.	-	Session facilitation	[Group interaction]
18.	-	Researching	-
19.	-	-	Booths
20.	-	Project samples	[Sample production]
21.	-	-	Training blocks
TRAINING STANDARDS			
22.	Taste discrimination	-	-
23.	Time	Time	Time
24.	Product attributes	[Reference samples]	-

Combined concepts chart

NO.	EXPERT 1	EXPERT 2	EXPERT 3
25.	End of scale values	-	-
26.	Variation	[Consistency]	[Consistency, ANOVA]
27.	Standard deviation	Standard deviation	Standard deviation
28.	Mean	Mean	Mean
29.	F sets	-	-
30.	-	Reproducible	-
31.	-	Fully trained	[Competency]
32.	-	Range	-
33.	-	-	Test exposures

Combined concepts chart			
NO.	EXPERT 1	EXPERT 2	EXPERT 3
SELECTION ACTIVITIES			
34.	Interviewing	Interviewing	Interviewing
35.	Data collection	-	[Questionnaire]
36.	Decision to hire	[Removal]	-
37.	Screening tests	-	[Screening tasks]
38.	-	Advertise	Advertise
39.	-	Screening sessions	[Session]
40.	-	Identify basic tastes	Identify basic tastes
41.	-	Threshold testing	Threshold testing

Combined concepts chart

NO.	EXPERT 1	EXPERT 2	EXPERT 3
42.	-	Triangle tests	Triangle tests
43.	-	Descriptive work	[Description tests]
44.	-	Fun activities	-
45.	-	Personality assessment	[Personality work]
46.	-	Data analysis	-
47.	-	-	Difference testing
48.	-	-	Time
OTHER OBJECTS			
49.	Training environment	-	-

Combined concepts chart			
NO.	EXPERT 1	EXPERT 2	EXPERT 3
50.	Population size	Population size	Population size
51.	Group leader	Group leader	Group leader
52.	Pay	Pay	Pay
53.	-	Panel/product attributes	Panel/product attributes
54.	-	Commercial pressure	-
55.	-	Panel expense	-
56.	-	Sensitivity	-
57.	-	Compusense	-

- indicates an assumption of no equivalent object

[] indicates an object which is assumed by the knowledge engineer to be equivalent to those on the same row.

Appendix F

Transcript of Consensus Meeting

Transcript of consensus meeting

with knowledge engineer and three experts - 30 Jun 93

1
2 KE OK, so this is the last substantial meeting that we'll all have and it's also
3 significant in that it's the first time that we've all come together as a group
4 in the knowledge engineering process. We've met and talked about the
5 group dynamics of the project but this is the first time that we've talked
6 about the knowledge bases without saying to you please keep the
7 transcripts to yourself, and so we're at quite a significant junction. Just to
8 review where we've come from to here; the individual interviews first the
9 first one covering the context moving forward to the second interview
10 where we tried to home in on more detail regarding that critical decision
11 point of how and when do we decide that a panel's fully trained. From
12 there you will recall that I sat down with each one of you and said 'what
13 we're going to do now is break the transcript down into areas where you're
14 actually expressing knowledge, hard data about the process as opposed to
15 places where you're expressing opinions and things like that. In most cases
16 you then went through and did most of that yourselves. From there, I did
17 the same exercise and cross-checked that we agreed with what was facts
18 and what was opinion. Then I broke those excerpts down into meaningful
19 phrases, basically one action per line, making it more readable and easy to
20 track and that's what you've got a copy of at the moment. I've then gone
21 through that line by line and identified key words and phrases that I've
22 termed objects, things that stand alone and relationships between them. A
23 quick, general reading of the transcripts gave me the overall feeling that
24 people were talking about the stages of the process, about training
25 activities, about standards, that might be better termed something else like
26 evaluation, I don't know. The headings of the groups aren't important. My
27 reading of it indicated that there were some groupings but it's only my
28 reading. To organise things, if I saw what I thought was a concept I would
29 write it down under the grouping that I thought it belonged to, given the
30 context that the particular person was talking in. For instance two people
31 have talked about triangle tests. One has talked about it in relation to
32 selection activities and the other in relation to training activities. It's just
33 a question of where it popped up. And also I've written down the line
34 number occurrences so we can track back. On that list you can see the
35 lines on which the item occurred. From that we also get some idea of
36 emphasis by the number of repeat appearances we see. So I made up the
37 individual list and then a 'doubles chart' as you see there in front of you
38 with all three of you line by line, broken down by those three sections and
39 where the words are all the same, I've drawn them all on one line. Where
40 I can't find any equivalent I've just put a dash and where I think that what
41 you're saying means the same as something else, I've put it in square
42 brackets. So that's the process to date and before I can go any further, I
43 need to start to model the domain at a fairly high level which will probably
44 represent the end of the project. Before I do that it's important that you

45 agree with my interpretations of your transcripts and secondly we agree on
46 a shared vocabulary e.g. if the term is going to be 'word generation' or
47 'word development' or whatever, we all agree that we are talking about the
48 same thing. Hopefully we will be able to get rid of the square brackets and
49 it may be that I've simply misinterpreted or missed something in your
50 words. The purpose of this interview then, is to achieve consensus on a
51 vocabulary in much the same way as your do with your panels. At the end
52 I'd like to be able to go out and say 'yes, we all agree that the stages are
53 this.' So where we've got things like triangle testing coming up in several
54 areas, which area should it belong to, or is it correct in all of them? Are
55 there any questions about the aim?

56 Expert 1 So when you talk about groups you've got training, stages of training,
57 selection activities and others?

58 KE Yes

59 Expert 1 And you want us to agree on the vocabulary to be used for each of those.

60 KE Well they're only my assessments. If you can think of a better way of
61 breaking them down or a better way of describing them then that's fine. It
62 has been suggested by someone else that a better way to group them is in
63 terms of... remember those key tasks we talked about in the first interview
64 of interviewer, planner, selector, that sort of thing?

65 Expert 1 Yeah.

66 KE Maybe it should be grouped around those groups. It's purely our decision.

67 Expert 1 No I don't have a problem about these groupings. I'm just a bit concerned
68 about training standards. I don't quite know what that term means.

69 Expert 3 I can't see that on mine.

70 Expert 1 Well it's on mine, on page ...

71 KE Well perhaps we can just talk about the actual names of these headings
72 first and then we'll come back to that expert 1 because I think it is an
73 important one. Stages of training on page 1 I saw as the contextual one,
74 the big one. There was a selection phase, a training phase, a project work
75 phase and then I came up with the decision point phase which seemed to
76 stand alone as it seemed like a discreet unit after training but before
77 project work.

78 Expert 1 Yes.

79 Expert 2 Yes that's right.

80 Expert 1 And sometimes you keep coming back to it later on.

81 Expert 2 Like it loops into all three of those. Like during selection you've got to
82 make a decision, in training there's a whole bunch of them and then again
83 in project work.

84 KE Yes. So is it fair to leave it as a discreet group.

85 Expert 1 Mmm. That was the thing I was most concerned about when we first spoke
86 is that the decisions we make are different between the three of us.

87 KE Ok. Stages of training we'll just leave as it is if that's the best description
88 for it.

89 Expert 1 Yeah I think that's pretty good.

90 Expert 3 Is project work still in the training phase or is that more when it's
91 completed?

92 Expert 2 That's kind of like the target.

93 KE Yes.

94 Expert 2 Cos I can remember you saying something to me when the project was
95 getting bigger and bigger that we were just going to focus on the point up
96 to where the panel was trained. And we stopped talking about project
97 work because to me that's the target of training.

98 KE Yes as opposed to decisions about training.

99 Expert 3 But the decision point... perhaps there's quite a few. There's one decision
100 point where you say that the panel is now trained and that's the finale.

101 Expert 1 But isn't that... like he's called it project work but should it be something
102 like end of training or something?

103 Expert 2 Yeah, or availability for project work

104 Expert 3 Or graduation to project work?

105 KE I think the reason that it came up was that some of you talked about using
106 project samples inside your training.

107 Expert 1 Yeah we did.

108 Expert 3 See in the retraining, that's in the ongoing monitoring

109 Expert 2 But also near the end of training

10 Expert 3 Mmm.

11 KE So what would you like to do with the overall group heading. At the
12 moment it's stages of training. Is there a better way of describing these
13 contextual areas?

14 Expert 3 If we're making a decision as to how to train a panel and not how to
15 monitor a panel or...

16 Expert 1 If it's just the training aspect I reckon the project work should be just end
17 of training, training complete or something like that.

18 Expert 3 But if we're also looking at the lifecycle of the panel over 2 or 3 years we
19 should also put in ongoing training. That would be another part of it.

20 Expert 2 Yeah, there's training to get the panel ready for project work but there's
21 also ongoing maintenance training.

22 Expert 1 So it's up to you how big you want this to be and expert 2 was right. I'm
23 remember you saying that this is going to get out of hand, let's just talk
24 about this perspective of it.

25 Expert 2 In which case project work should be like the graduation sort of thing.

26 KE OK, what word best describes it then?

27 Expert 3 What did you say expert 1?

28 Expert 1 I said training complete

29 KE OK so we had a selection phase, then we had a decision point, sequentially
30 that is.

1 Expert 2 And that's really it.

2 KE So if we take our objective as 'when are they trained' are you happy to
3 stop at that point?

4 Expert 1 Yeah

5 KE OK so we'll delete line 3 there.

6 Expert 2 Like that decision point to me has two outcomes.

7 Expert 1 Yes they're ready, no they're not, let's go back.

8 KE OK. These are in no particular sequence. Training activities is the second
9 group and this just seems to me to be the group of things that happened

40 after you'd got them off the selection activities and... I'm just thinking
41 maybe we should do this in sequential order. Shall we jump over into
42 selection activities?

43 Expert 2 Mmm
44 KE How suitable does selection activities seem as a heading for that group?

45 Expert 1 Yeah

46 KE Following on, just looking at the group headings first. After selection
47 activities we have training activities.

48 Expert 3 Yeah

49 Expert 1 Mmm

50 KE OK and then training standards which may or may not be the correct
51 heading. I know that you've got some thoughts on it expert 1, what do you
52 think?

53 Expert 1 Looking at your list there, they're more training attributes, I think
54 standards imply something to us that...

55 Expert 2 I see what you mean about it applying to us but I see... some of my
56 answers there were when you asked me how I measure my training.

57 Expert 1 Oh OK

58 Expert 2 So all those; mean standard deviation etc...

59 Expert 3 Competency standards or something like that?

60 Expert 1 Yeah

61 Expert 2 Performance standards perhaps?

62 KE I'll tell you where I ran into some difficulties because I know that you all
63 talked about using standard deviation and the like in your selection
64 activities as well as in your training activities. We're limiting our view to
65 training activities but some of these apply, like reproducible etc to select
66 as well.

67 Expert 1 So if we use something a bit more vague like performance measures, that
68 could apply to selection and training and... could it?

69 Expert 2 Mmm and then there is probably other little things we could group in
70 there...

71 KE OK, so are you happy with performance measures?

72 All Yeah.

73 Expert 1 So where would that fit then in your order?

74 KE Well this order may change. It's just a building block for us to work with.

75 Expert 1 Oh OK

76 KE We'll end up with a lot of flowchart like diagrams which will be models,
77 selection models, training models, decision making models. There'll be a
78 sequence but it may jump between them as we currently have them. All
79 right, well leaving performance measures then, I ran into a whole lot of
30 things that didn't seem to fit anywhere else.

31 Expert 2 Other. Yes we do that too.

32 KE You know things like the size of the group, group leader attributes and so
33 on, pay, remuneration, commercial pressure and expense. Things like that.
34 The training environment which loosely relates to the round table room
35 etc.

36 Expert 2 What did we refer to the control issue as?

37 KE You mean like the parametric problems of size of group, too many or too
38 few?

39 All Mmm.

40 KE Do you remember at the beginning expert 1, in the scene setting, you
41 talked about the need to have something that you could hand a new
42 person on joining the unit. You could say "you are going to be a panel
43 leader, read through this and there are things listed that will act as a guide
44 for you..."

45 Expert 1 Yeah like if you're training a hundred people...

46 KE So that's why I noted them down.

47 Expert 1 Yeah well could you say something like guidelines?

48 Expert 2 I was going to say external factors but then they're not all external.

49 Expert 3 Supporting guidelines?

0 KE Are they panel dynamics you know the type of people, where they are and
1 who is leading them.

02 Expert 1 How would that affect the training environment though? Like the physical
03 stuff. Is that part of panel dynamics?

04 Expert 3 What do we actually mean by training environment?

05 Expert 1 I'm not actually sure.

06 KE Training environment , as I recall, and it was on line...

07 Expert 1 Line 6 I think.

08 Expert 3 That would be something like if we went out to a factory to train a panel
09 wouldn't it?

10 KE Yes. The other thing that occurred to me with regard to training
11 environment is having them round the table. I'm thinking that maybe I
12 should have grouped those because in training, we find... see expert 3
13 mentioned booths under training activities. Now in some ways that strikes
14 me as a similar context to the training environment.

15 Expert 1 I notice that expert 2 and 3 have talked about...a couple of lines I just read
16 then...are referring to the type of panel eg product 1 or 2.

17 KE Yes and I've mentioned panel attributes in one place somewhere. The
18 group of words that I used to try and bring it all together, eg expert 2
19 listed panel attributes and expert 3 said panel/product attributes and I've
20 put them on one line. That's what I was trying to say.

21 Expert 2 Yeah I can remember talking to you about a whole lot of the actual
22 decisions which relate to the product whether it was complex or not. But
23 as to what to call all these things.

24 Expert 1 It's not extraneous factors is it? No. What about process influences or...I
25 mean what's the big word you going to use for this thing. Is it a process
26 or...

27 KE Well it will be what you want and you said that a prescriptive model was
28 closer to what you wanted, but it was hard because we agreed that there
29 would probably never be one answer but a prescriptive model was going
30 to be more useful. I don't think that we're going to be able to get down to
31 hard numbers like so many tests etc.

32 Expert 1 So if we said model features or model...

33 Expert 3 What about influences?

34 KE Yeah

35 Expert 2 Considerations

36 Expert 1 Yeah, other considerations.

37 KE OK

38 Expert 2 Yeah considerations is cool.

39 KE OK so we'll make one of our groupings 'other considerations'

40 Expert 1 Yeah it's a matter of semantics but I can see what you're trying to do, to
41 get us all to acknowledge what we do is the same.

42 KE Yes. At the end of the day you see, if you ended up with a knowledge base
43 founded on a vocabulary that you didn't all agree on it wouldn't function.

44 Expert 1 So what you'll end up with here is once you've clarified the other things,
45 everything else will go under 'other considerations' isn't it. Anything that
46 doesn't fit in any of those other ones.

47 KE Yes but at the end of the day that might not be too significant when we
48 build the models because the models might draw on all the groups with
49 just a logical connection between attributes.

50 Expert 1 Right

51 KE It's a transitional step. But it is important that we all achieve agreement
52 along the way. OK can I just then flash back to the first part, the stages of
53 training bit. I think that we agreed there that we just have the three lines
54 there, selection, training and the decision point, for the contextual phases.
55 Is that right?

56 All Mmm.

57 KE Is there anything that you want to add having re-read your transcripts.

58 Expert 2 To there or for the whole thing?

59 KE Just to that point. Are there any other important stages that you feel have
60 been missed.

61 Expert 1 It doesn't sound exciting but that's basically it.

62 KE This is something that's interesting, not just with this process but with
63 performance processes generally is that some people are very disconcerted
64 to find that their previously extremely complex job has been reduced to 5
65 objectives or so.

56 Expert 1 Yeah. You're tearing us apart on this one you know.

57 KE Well it's good because you can now focus on performance. You're not
58 cluttered with all this other baggage.

59 Expert 1 We had managed to create a mystique around this job you know.

70 Expert 3 Yes this could ruin everyone's impression...

71 Expert 2 Yes people will say "they only do three things..."

72 KE OK. Now we're getting further down into the nitty gritty of it. Looking at
73 selection activities was the next thing wasn't it?

74 Expert 1 Yeah.

75 Expert 2 There are some things here that we've done more recently...

76 KE OK

77 Expert 2 Specific things like ranking tests

78 Expert 3 Is that not down there?

79 Expert 2 No, much to my surprise.

80 KE OK.

81 Expert 1 Yeah I guess my areas were more vague where you guys went into the
82 detail

83 KE Yes, you talked about screening tests which is a fairly broad topic. That's
84 the intuator part of you. Big picture stuff.

85 Expert 1 But I don't want to be one of those!

86 Expert 2 I keep telling you expert 1.

87 KE Expert 2 and 3 had screening tasks too but when we went on down we
88 found more detailed descriptions like identify basic tastes, thresholds,
89 flavours, descriptive work, fun activities, personality assessment and
90 difference testing. They were the real nitty gritty ones so is rank ordering
91 a different activity again?

2 Expert 2 Well, you could just write discriminatory tests and that would include a
3 whole range of techniques or you could list out the different techniques
4 that come under discriminatory tests.

95 Expert 1 If we were a bit more vague like discriminatory tests, that means that this
96 model will hold for a lot longer because they'll keep inventing new types
97 of tests that will fit in under either analytical or discriminatory, wouldn't
98 they.

99 Expert 2 Mmm.

00 Expert 1 It's a shame that the recorder can't pick up sign language there.

01 KE They do this on video cameras as well to capture things like that expert 2.

02 Expert 2 You're supposed to say could you give me an example...

03 All ...laughter...

04 Expert 2 Instead of listing out things like triangle tests, ranking etc, could we just
05 call them discriminatory tests?

06 Expert 3 Yeah because that's what they're often referred to as.

07 Expert 2 Yes and that would give the model more flexibility. If new discriminatory
08 tests were identified they could be included without disruption.

09 KE OK so are we going to go with that one?

10 Expert 1 Yeah I'd be quite happy with that

11 KE And what is that concept going to replace in terms of..

12 Expert 1 They will take triangle tests and difference testing...

13 Expert 2 Yeah and what I was saying about ranking...

14 Expert 3 Probably threshold testing

15 Expert 2 Is it

16 Expert 1 Don't know expert 3.

17 Expert 2 I personally think that should be identified in a separate category

18 Expert 1 Yeah I agree

19 Expert 1 So you had threshold, discriminatory, description

20 Expert 3 Descriptive covered...

21 Expert 1 Yeah that would cover it

22 KE What was the analytical ones you mentioned earlier.

23 Expert 1 Oh that's sort of descriptive but analytical is a little bit different.

24 KE So all three of you as I read it are happy with discriminatory...

25 All Yeah

26 KE ... as a selection activity; discriminatory tests and that replaces triangle
27 tests, ranking and difference testing?

28 All Yeah

29 KE OK and threshold testing is kept separate? Expert 1 are you buying into
30 threshold testing as a separate entity

31 Expert 3 And that probably covers identification of basic tastes, so that's redundant

32 KE So threshold testing gets removed

33 All No no

34 Expert 3 Well remember that we can just do threshold testing for aroma, you know
35 what I mean?

36 Expert 2 Yeah true.

37 Expert 3 You know the threshold at which they perceive an aroma

38 Expert 1 Yeah, when we say identify basic tastes when can do that through
39 threshold testing, or through ranking or through other things

40 Expert 3 Yeah

41 Expert 1 So I think it's covered in...

42 KE Identify basic tastes? are you happy with that expert 2 and 3?

43 Expert 2 Mmm

44 Expert 3 I think...

45 Expert 1 Do you think it should still be there?

46 Expert 3 No no

47 Expert 2 I wonder what I meant about removal?

48 Expert 3 Have a look on the line number in you transcript to pick up the context.
49 That's what I was doing.

50 KE OK shall we try and get a sequence as we go along? What's the first thing
51 that you do on the timeline in selection activities?

52 Expert 2 Advertise

53 KE Are you happy to include advertise in your list expert 1?

54 Expert 1 Yeah

55 KE Second?

56 Expert 3 Umm, well I put down questionnaire because we get them to fill out that
57 pre-screening questionnaire, remember I did that for one?

58 KE The data collection that you referred to expert 1 was over the phone, who
59 are you, how are you

50 Expert 1 Yeah that's what expert 3 is talking about

51 KE That's what I figured. That the information is just gained in a different way

52 Expert 1 Yeah. I've called it data collection. Should it be questionnaire expert 3?

53 Expert 2 Which questionnaire was that?

54 Expert 3 Well we have that pre-screening questionnaire on the phone...

55 Expert 1 Yeah, this is you name, address etc...

56 Expert 2 Oh OK sorry!

57 Expert 3 That's data collection probably. I have a pre-screening questionnaire as
58 well where you ask them like "do you wear false teeth?" all that sort of
59 thing if they're going to be on the texture panel. So I reckon that data
60 collection is demographics and then pre-screening questionnaire is
61 different.

72 KE You two happy with that?

73 Expert 1 Yeah you're right, I agree

74 Expert 2 Yes you can't just say do the screening tests

75 KE Yes - screen the applicants - end

76 All ...Laughter...

77 KE So are we going to end up with two? We've got data collection which is
78 just demographics is it?

79 All Yes

80 KE Are you all happy to include that in your lists as one of the selection
81 activities?

82 All Yes

83 KE And then we've got the questionnaire which you called what expert 3?

84 Expert 3 The pre-screening questionnaire

85 KE And you two are happy with that?

86 Expert 1&2 Yes

87 KE And data collection is the second in sequence?

88 Expert 1 Yeah

89 KE And then what's third?

90 Expert 1 Pre-screening questionnaire, then screening tests

91 Expert 2 Yes screening tests, any of those three types we discussed earlier

92 Expert 3 All of those three types

93 Expert 1 Probably basic tastes first

94 KE The category that you've got on line 37 then includes discriminatory tests,
95 threshold tests and descriptive work and those three are subordinate to...

96 Expert 2 And probably threshold testing would probably be done first

97 Expert 1 Yeah in order it would probably be threshold, discriminatory, descriptive

98 Expert 3 And we repeat it. Do you have to have that down?

99 Expert 2 Mixed up with fun activities, personality assessment...

100 Expert 3 What about degree of difficulty because they get harder as the week goes
101 on

02 Expert 1 We haven't mentioned that expert 2

03 Expert 2 No we haven't

04 KE Well we've got the activities there. Maybe when we get over to
05 performance measures we can have degree of difficulty there somewhere.
06 OK so where have we got to. Personality assessment and personality work
07 are listed. What's the one term that you want to use to cover that side of
08 things in selection, the people side?

09 Expert 1 Personality evaluation?

10 Expert 3 Yeah personality evaluation or...

11 Expert 2 Screening out the undesirables?

12 Expert 1 It's a bit of a numbers game though

13 Expert 2 Yeah like what expert 1 is talking about is if we get to the end of the week
14 and we think gosh those tests were too hard we might lower our standards
15 or we might have someone whose scores are a little bit lower than we'd
16 like but from the personality attributes, they'd be great to have on the
17 panel.

18 Expert 1 But you're right, there is a subjective/objective component mix

19 Expert 3 Yeah there is a bit of that

20 Expert 2 Like ones that do really well but I won't have them on my panel

21 Expert 3 Cos that is the whole picture, those three things

22 Expert 1 And I mean it gets very fuzzy if the KE is trying to develop a flowchart
23 that everyone follows and they get to this grey area

24 Expert 2 But isn't it a part of performance evaluation though because there is their
25 score and the suitability of that score and so deciding on people on the
26 basis of their personality is part of that and our individual observations

27 Expert 1 Yeah well one of the concerns that I have is that I might choose someone
28 different to you

29 Expert 2 Well that's OK because you've got to work with them

30 Expert 1 That's true

31 Expert 2 It would only become a problem if you took over my panel

32 KE So it's quite a complex interaction of things

33 Expert 1 Yes it is

34 KE At the end of the day then do you want to leave that decision to hire as a
35 separate issue to the performance evaluation or do you just want to put
36 them together?

37 Expert 2 You weren't thinking expert 1 of the decision to go ahead and put the
38 panel together were you

39 Expert 1 No I can see what KE is getting at. We evaluate their performance and
40 we've got all of the lists of their triangle tests and how they did in the
41 interview but at the end of the day we still have to make a decision

42 Expert 2 Could we refer to it as personnel evaluation and selection?

43 Expert 1 I just think that there's one more step, decision to hire, just like...

44 Expert 2 No that's fine

45 Expert 1 I think expert 3 said it, someone said it

46 Expert 2 You said it

47 Expert 1 Me! Oh I did too, sorry about that

48 KE OK, shall we go with that

49 All Yes

50 KE OK moving now from selection activities to training activities. Straight away
51 we see threshold tests and triangle tests pop up again which we've just
52 grouped in the selection activities. What do you want to do with them this
53 time.

54 Expert 1 As long as it's covered somewhere else I'm happy

55 Expert 3 I still think that they're part of the training activities

56 Expert 2 I think that we should just call them discriminatory tests because lately I've
57 used ranking

58 KE OK what's the first thing you do when you start training. Let's try and get
59 a sequence

60 Expert 1 Well it depends. The first thing you do is introduce the panel to each
61 other. I mean it's a general introduction

62 KE OK so on line 10 we've got group dynamics

63 Expert 2 It's more than that, like there's administration stuff, confidentiality
64 agreements, this is what you're here for, this is what establishment 1 is
65 about, this is how to taste, this is what taste buds are, you know a general
66 education type thing

67 Expert 3 Yeah that's normally in the first session

68 KE What will we call that

69 Expert 3 Administration?

70 KE Two of you referred to panel protocols and that's what I grouped those
71 things under, how to taste, how to spit politely etc

72 Expert 2 It sort of is panel protocols isn't it that's quite right but there is also
73 administration. Maybe there's two

74 KE OK two of you mentioned that but expert 1, you didn't go into that sort of
75 detail. How happy are you with administration and panel protocols?

76 Expert 1 Yeah I think that's good

77 Expert 2 And there's the fuzzy stuff like introductions and little games

78 KE OK does that refer to line 17, session facilitation and group interaction. I
79 can recall that these were ongoing activities. How happy would you be with
80 starting at the first session accepting the fact that it carries on right through

81 All Yeah

82 Expert 1 You can tell I've been out of this for a while, can't you

83 KE Do you want to call it session facilitation or group interaction or something
84 else

85 Expert 3 Facilitation is quite good because you are trying to facilitate

86 Expert 1 Yes

87 KE OK, are you happy with dropping group interaction expert 3

88 Expert 3 Yes

89 Expert 2 That's very noble expert 3

90 KE After that what happens?

91 Expert 3 Are we up to stage 3 now?

92 KE Training activities. We've had panel protocols, administration, session
93 facilitation...

94 Expert 1 Word generation

95 Expert 3 Yeah

96 Expert 2 Maybe review of basic tastes. We quite often do that on the first session

97 Expert 1 Isn't that cos we're desperate for things to do?

98 Expert 3 What about warm-up activities like basic tastes, difference testing.
99 Sometimes we do that sort of thing

00 Expert 2 Just to make them more confident

01 Expert 1 Yeah

02 Expert 3 Confidence builders

03 KE OK, so what are we going to call them?

04 Expert 1 Are we allowed to use non-technical language to describe them

05 KE Sure, they're your words, it's your knowledge!

06 Expert 1 How about warm-up activities?

07 Expert 2 Yeah they are really

08 KE OK all happy?

09 All Yes

10 KE OK then we move on to word generation, which I grouped with product
11 description. Is that the same thing?

12 Expert 2 Yes

13 Expert 3 Vocabulary generation?

14 Expert 2 And descriptive work?

15 Expert 1 Oh hell!

16 Expert 2 I'm just trying to think what word we normally use

17 Expert 1 Development of descriptors

18 Expert 2 Yeah

19 Expert 3 Vocabulary development

20 KE OK we have three different options. What's it going to be

21 Expert 3 Vocabulary development

22 Expert 2 I personally think attributes come into it but I won't say that.

23 Expert 3 Actually in some ways vocabulary development is the whole thing because
24 what you do is start off with word generation and when you've got all those
25 words you develop a vocabulary

26 Expert 2 It's more development than generation

27 KE So could we agree on the word development?

28 Expert 3 I actually disagree. I think that in the first instance it's generation and later
29 you develop it into a vocabulary

30 Expert 1 Yeah. Keep drinking your tea KE, I apologise for this unruly group (all
31 laugh)

32 Expert 3 No no, it's not like that

33 Expert 2 I agree, it's word generation

34 Expert 3 Descriptor generation?

35 Expert 2 Yeah, descriptor generation and then descriptor development

36 KE So you want to have two different concepts there?

37 Expert 2 Yeah

38 Expert 3 Yes because at the descriptor development stage we go away and try and
39 make up things that symbolise those words

40 Expert 2 And that ties on to reference sample development

41 KE OK

42 Expert 3 Well we could have reference sample development and training sample
43 development meaning those used later on

44 KE OK so we've got descriptor generation and descriptor development and
45 then to where?

46 Expert 2 Reference sample development and training sample development

47 Expert 1 Well wait a minute. Are we doing words to numbers first or what?

48 Expert 2 No often you don't because you've got to know what the end of the scale
49 is before you can

50 Expert 3 They've got to know what the word means before...

51 Expert 2 What we haven't got on here is the introduction of intensities. We've gone
52 words to numbers but we haven't gone to words. Word generation refers
53 to the descriptors but when we say words to numbers that refers to
54 intensity work

55 Expert 1 Mmm

56 Expert 3 Well I think that we should put down that we develop reference samples
57 and training samples at that stage and it's sort of a loop at that stage until
58 we get all the vocabulary sussed. Then we start to look at intensity of that
59 particular product by attribute

60 Expert 2 ... but say you're wanting to develop references for saltiness, you've got to
61 say well this is low on the reference scale and this is high on the reference
62 scale

63 Expert 3 Yeah. I'm just trying to think of training samples and just trying to get
64 them to identify that particular attribute is exemplified in that particular
65 product

66 KE OK. On line 14 where we've got scale development, where does that come
67 in? Because we started off getting end of scale points then you developed
68 your descriptor scales, line scales and it was a very broad term to
69 encompass all that discussion of different things, e.g. 150mm lines etc. Do
70 you want to keep that as a separate issue to what we're currently talking
71 about with development of samples?

72 Expert 2 To me that scale development is a continual part of all these other
73 developments. I mean we might go in with a fixed idea that we're going to
74 use this scale and so the learning of that scale is integral to the learning of
75 the reference samples. Anyone disagree?

76 Expert 1 Hold on, I'm just trying to think, before KE says about his flowchart; can
77 we have a flowchart that comes straight down and then goes off to one
78 side that says scale development, reference development, training sample
79 development etc but in the main model it's just referred to once

80 KE Sure. What you'll find is that they're not flowcharts in the sense that you
81 understand that term. They're models that come out of a library of task
82 models anyway so don't worry, I'll find or amend a model that fits what
83 you're doing

84 Expert 1 That's alright then

85 Expert 2 Yeah it is circular

86 KE You see we may plug a decision making model into that loop which may
87 become a separate aspect of the model on its own

88 Expert 1 That's alright then

89 Expert 3 Well perhaps it should go vocabulary development, training sample
90 development then reference sample development, interpretation of the
91 scale, introduction of the scale

92 Expert 2 Cos you normally work through a whole bunch of samples before you find
93 one that everyone agrees on

94 KE So that's developing your training sample is it?

95 All Yes

96 Expert 2 So that stops being just a sample that you made up on the day for training
97 but becomes one that you can refer back to for a reference, and the panel
98 learns that

99 Expert 1 Yeah so training comes before the reference

00 KE And so training sample development?

01 Expert 2 Yes

02 Expert 3 And then reference sample development

03 Expert 1 Have you ever tries to do these things with more than three experts KE?

04 All Laughter

05 KE OK reference sample development and then to where? Are we talking
06 about scales then?

07 Expert 3 Well scales are sort of part of that like you'd say this is a reference sample.
08 Where would you scale it?

09 Expert 1 So they've got to know what the scale is? Well we've got training samples

610 before reference ones. Part of being a reference is that you found it to be
 611 an end of the scale sample. So can you have training sample development
 612 like you've got there, then introduction to scales, then reference sample
 613 development

614 Expert 2 Yep you can like if I get lots of caramel words, I'll get a sample which I
 615 think is caramel and I'll say "is this what you call caramel" and if they say
 616 yes I'll say "fine, we'll use it as a reference to develop the scale from"

617 KE Sorry, just to go back then, does scale development begin before training
 618 sample development or...

619 Expert 2 No, between training and reference

620 KE OK, got it. OK and then to where?

621 Expert 2 Well feedback and analysis of performance are an ongoing cycle, especially
 622 once you get on to the scale that you're going to use

623 Expert 3 So now I'd say feedback and analysis

624 KE Yes as you can see, you all mentioned the same terms here

625 Expert 2 Yes we did quite well at that didn't we

626 KE And they start from what point, from the reference sample development?

627 Expert 2 I think it starts right at the beginning, just positive reinforcement stuff for
 628 the panel, even during word generation

629 KE At the start of the panel protocols or the warm-up activities or the
 630 descriptor development or where would you like to start that off?

631 Expert 2 Descriptor development?

632 Expert 3 It depends on what we're defining feedback as. Cos I mean feedback could
 633 mean "oh yeah you did really well"

634 KE Feedback came up in many different contexts. It included just building up
 635 their confidence and also supplying them with their standard deviations etc.
 636 Can we just accept that it's an ongoing part of the whole process?

637 All Yes

638 Expert 2 Certainly there's an element of it in performance measures

639 KE OK. Are you happy to move it to there?

40 Expert 1 I might argue with that one

41 Expert 3 Yeah you need to analyse performance before you can move on to the
42 next stage

43 Expert 1 I think if we don't have it in this step...

44 KE These aren't linear remember they do overlap. One is a concept and the
45 other one a relationship

46 Expert 3 Well I suppose analysis of performance is part of performance measures

47 Expert 2 Yeah you can't go from training samples to scales to reference samples
48 without measuring performance

49 KE Yes. If you liken analysis of performance to a door. Once you're through
50 it opens the way to using all those other measures

51 Expert 1 Well we're talking about activities in training and analysis of performance
52 is an activity that you do in training. Feedback is an activity that you do in
53 training. Whether it's one before the other is ...

54 KE Shall we leave it as ongoing activity?

55 All Yes

56 KE OK now moving, expert 1 you mentioned product evaluation and I put it
57 on the same line as reference samples

58 Expert 1 Yeah that's fine

59 KE Good

60 Expert 2 On the next thing on page 18 where I've got researching

61 Expert 3 What does that mean?

62 Expert 2 That kind of comes under training samples

63 Expert 1 Oh OK

64 Expert 2 How to make them, looking at the literature and so on

65 KE Yes you talked about chemical reactions and things

66 Expert 2 Yes that comes into that whole area of sample development

67 KE OK so would you drop researching as a separate entity?

68 Expert 1 What's training blocks expert 3? Was that mental blocks to...

69 KE How do you feel about the researching issue?

70 Expert 2 It's not a sequential activity I don't think but it can be a part of all the
71 sample areas

72 KE OK are you happy to leave it as part of them or would you like it to be
73 listed separately?

74 Expert 2 Well I think it can be quite important

75 Expert 3 It is important. I mean someone who comes in who hasn't trained a panel
76 before is going to have to go up to the library and see what are the
77 important factors in that product

78 Expert 2 Yeah it probably comes in there but it also comes into like choosing the
79 scale

80 Expert 1 Yeah but it also comes right at the beginning before that lot

81 Expert 3 What comes right at the beginning, just groundwork

82 Expert 2 That's something that we haven't even got on here which is planning

83 Expert 1 Yes planning

84 Expert 2 But also literature review

85 Expert 2 Yeah and methodology, like how are you going to present the samples,
86 which order and so on

87 Expert 3 Oh yeah

88 KE Would you like to mention planning as a training activity. It's not there and
89 it brings together researching and some of these other issues

90 All Yes

91 KE Or is planning a much higher level activity?

92 Expert 3 Well maybe it is because before we do screening, we have to plan what
93 screening we do

94 KE Maybe it should be with the big three or four at the front?

95 Expert 3 Yes maybe it should be one of the stages of training. Is that what you're
96 saying?

97 Expert 1 Yeah

98 Expert 2 Yeah in the first four lines it should include planning

99 Expert 3 Yes and it should be the first one

00 Expert 1 Yes

01 KE OK all happy with that?

02 All Yeah

03 KE OK so does that then remove researching from the training activities list?

04 Expert 2 Well it's mostly in the planning but there is some that's ongoing

05 Expert 1 Well I imagine that KE wants us to list under planning now that we've
06 chosen it the activities

07 KE No not necessarily. Not at this stage anyway

08 Expert 1 And that would include literature searches, methodology, appropriate
09 selection of samples etc

10 KE At this stage what it does is alerts me to the fact that there's another part
11 of the model about to be built. If you imagine breaking this into chunks.
12 The four we've got now are chunks.

13 Expert 3 I think the researching should come in at training samples

14 Expert 2 Yes if there is a separate thing for researching it is at the training sample
15 time

16 Expert 1 I don't think it's a very good word though

17 Expert 3 What about lab work?

18 Expert 1 I mean what are you trying to research again. Run this by me again

19 Expert 2 You're trying to research if anyone's done it before and what they used.
20 You may be trying to establish the different methods of producing burnt
21 flavour say.

22 Expert 3 Well maybe we could leave it out and just... When we defined training
23 sample development we could put it in that. Will we be further defining
24 that?

25 KE We may further define some but not all of these

726 Expert 3 So training samples could be further defined by...

727 Expert 1 Yeah

728 Expert 2 OK that's OK

729 KE Ok, what are you going to do. What's the consensus?

730 All Leave it out

731 Expert 3 That training blocks thing can get left out. What I meant by that is when
732 you do say two weeks on vanilla flavour then you go to oxidised flavour

733 Expert 2 Oh I see what you mean

734 KE OK is that an important thing to focus on. Relating it in those terms, I
735 wonder if others mentioned it in other ways

736 Expert 2 Because you do try and have a concerted effort on something then you
737 might give up for a while and then come back to it. I don't know, it just
738 might require a better definition

739 Expert 1 I mean you'd do that in the performance evaluation think "this is a waste
740 of time, I'll go and do something else"

741 Expert 2 Or "this is going really well, I'm on to something here"

742 KE Is it a loop in your decision-making cycle rather than a discreet object on
743 its own?

744 All Yes it is

745 KE Project samples, sample production where does that fit in on line 20?

746 Expert 2 Back at training samples

747 KE OK so should we just integrate the training sample development

748 Expert 2 It actually comes in before and after

749 Expert 3 Actually, I think that is part of planning, that particular bit. Because you
750 look at your sample production. You look at the project samples you've
751 got coming up

752 Expert 2 Yes and what kinds of things the panels likely to have to know about

753 KE So shall we note it as part of the planning?

754 All Yeah

755 KE One that I missed was the words to numbers thing. Where did I put that?

756 Expert 1 That's when we were talking about introduction of scales

757 KE Should it just be left as part of that scale development thing?

758 All Yes

759 KE OK how about this threshold testing, triangle tests. Expert 1 you mentioned
760 them. Are they covered elsewhere?

761 Expert 1 That's covered

762 KE Covered by what?

763 Expert 1 Covered by earlier on we said there was introduction, didn't we?

764 Expert 2 Also that's a task you might do under training sample development

765 Expert 1 Yeah

766 KE Both of them?

767 All Yeah

768 Expert 2 And I think that should be just discriminatory tests rather than just those
769 two things

770 KE Just discriminatory or do you want to use those other categories...

771 Expert 3 No just discriminatory

772 Expert 2 But imagine someone coming in and reading this. Shouldn't we indicate
773 that some tests are descriptive and that there is vocabulary development

774 Expert 3 Yeah it should be

775 Expert 2 What do you think KE? Because basically all of panel training is
776 descriptive really

777 KE Yes. Well I'm getting a better feel for it as we go on. I didn't realise until
778 now that there are descriptive and discriminatory divisions before. It was
779 just a string of activities

80 Expert 1 Right

'81 KE So as I understand it then, then individual tasks are being subordinated to
'82 sample development, reference and training samples and we've got
'83 discriminatory tests although threshold tests might be separate to that

'84 Expert 3 Mmm. Sometimes we test their threshold on a particular attribute

'85 KE OK. Happy with that?

'86 Expert 1 When we started that we came up with introduction and something or
'87 other. And in those warm-up activities, that's where I would do it

'88 Expert 3 Oh well I also use it to see say with oxidised flavour and milk fat to see if
'89 they can pick up a difference if they haven't used that word to describe the
'90 samples because they may not be perceiving it at all

'91 Expert 1 Oh that's fair enough

'92 Expert 2 You're thinking of threshold testing of basic tastes

'93 Expert 1 Yeah

'94 KE You'll see on line 11 'training session'. Two of you mentioned that and one
'95 didn't specifically. The reason why I put it there was that it gave a handle
'96 on duration and frequency of training. What are your thoughts?

'97 Expert 3 What, in terms of leaving it in there as a performance measure? as a
'98 training activity?

'99 KE It's the vehicle by which it's all done

'00 Expert 3 Yeah it is really

'01 Expert 1 Wouldn't it become part of whatever that other word is we're going to
'02 use?

'03 KE Can do. You might recall that I asked you all a lot of questions on why you
'04 run 3 sessions per week because that was one of the few hard data sources
'05 mentioned

'06 Expert 3 Yeah it is partly other considerations though

'07 Expert 1 Yeah cos it would be different if you asked Lyn, for example, a woman
'08 that is visiting us; they don't run their training sessions the way that we run
'09 ours. They run theirs all in one day

'10 Expert 2 Yes or two on Thursday and two on Friday for two and a half hours each
'11 time

312 Expert 1 So I think it would be in that other area

313 KE Other considerations?

314 All Yes

315 KE OK so training sessions have gone to the other considerations area. Now
316 group co-ordination and group dynamics also came up there when we were
317 talking about training activities

318 Expert 3 That's more an ongoing thing.

319 KE So you don't do any specific group related activities in your training?

320 Expert 1 Not really

321 Expert 2 Not unless something starts going terribly wrong

322 KE So should we delete line 10 then?

323 All Yes

324 Expert 2 And we haven't gone on about half-blinds and booths and so on

325 Expert 1 Well I have. I wondered when he was going to bring it up

326 KE On line 13 expert 1 mentioned half-blind tests

327 Expert 2 Well done expert 1!

328 Expert 1 See I do remember some things

329 KE So what's the situation there?

330 Expert 1 That comes into the analysis of performance

331 Expert 2 Yes like 'we're going to try this technique now because we want to see
332 whether they can do it on there own.' So it's not always at the end

333 KE Does it relate to the use of the booths?

334 Expert 2 Yes

335 KE Yes expert 3 mentioned booths. Should that be on the same line as half-
336 blind tests?

337 All Yes

38 Expert 1 I don't like that word but I can't think of any other way of saying it

39 Expert 2 To me the reason you do it is a measure of individual performance

40 Expert 1 That is a much better way of talking about it

41 Expert 2 I mean normally if you're in the middle of training and you stick the panel
42 in the booths it's because you want to check out whether certain individuals
43 can do it on their own or not. So it does come under that performance
44 measurement

45 Expert 1 Well should we have performance measurement as a group and as
46 individuals then?

47 Expert 3 Yes

48 KE OK so line 12 'analysis of performance"; should we make that 'analysis of
49 group performance' and then on line 13 make that 'analysis of individual
50 performance. Happy with that?

51 All Yes

52 KE And would that then be acceptable to drop off booths and incorporate that
53 with analysis?

54 All Yes

55 KE OK. Performance measures then. Putting aside the actual numbers then,
56 different people brought up different aspects of the measuring area and
57 you can see them there. Some came through all the time like standard
58 deviation and mean. I don't see that there is a sequencing problem in
59 performance measures. You just use what's appropriate as a grab bag of
60 measuring instruments

61 All Yes

62 KE So we are we all happy with standard deviation and mean. You all said it

63 All Yes

64 KE Perhaps then we can just work our way down from the top. On line 22 you
65 mention taste discrimination as a measure expert 1.

66 Expert 1 Taste discrimination. That's basic taste discrimination

67 Expert 3 Yeah I'd agree with that as a measure. We have to see whether they can
68 distinguish the taste

69 KE OK so you're happy with that?

70 All Yes

71 KE OK, time

72 Expert 2 What did you mean?

73 Expert 1 All of us said it!

74 KE I recall it as a measure of how long it took to run a panel, it was a
75 measure of all sorts of different things

76 Expert 2 So do you mean like how long a panel takes to train like 3 months or 6
77 months?

78 KE Yes

79 Expert 1 Yeah I was thinking also of the time to run selection like 1 versus 2 weeks

80 KE Yes, it's an attribute that will pop up in lots of different places

81 Expert 1 Put it this way. If I saw one of my staff spending 4 weeks on screening, that
82 would be a measure of performance to me

83 Expert 2 And I remember saying that if I had been training a product panel for 4
84 months then that would be a measure of my poor performance

85 Expert 1 Yes don't remind me of that one expert 2

86 KE OK we'll leave it there. Now line 24, I'm not really sure how to interpret.
87 Product attribute, samples, we really have taken those back into training
88 activities

89 Expert 2 Well what I meant was that you can use them as a performance measure
90 by giving them to them blind and comparing them to what they say they
91 think, using duplicates and the like

92 KE And they should be able to spot the duplicates?

93 Expert 1 Yeah. I looked at it from the point of view that some product attributes
94 are a lot harder than others.

95 Expert 2 Yeah. We didn't mean the same thing

96 KE Right. What are we going to do with them then

97 Expert 3 What is that called. There's a word coming out. Reference samples?

98 Expert 2 Well to me it's a measure of how much they've actually learnt

99 Expert 3 Yeah but there's a word for that, putting in reference samples that they
00 don't know about

01 Expert 1 Blind samples, blind reference

02 Expert 3 Unknown reference samples?

03 Expert 2 Yes, like I give my panel a reference sample and see where they put it
04 compared to where they say they'd put it

05 KE Yes. So what would you like to call that process?

06 Expert 1 Well could you say 'use of blind reference?'

07 KE Blind references?

08 All Yeah

09 KE Are you all happy with that replacing reference samples and product
10 attributes?

11 Expert 1 No it wouldn't replace product attributes

12 KE OK then. Blind references

13 Expert 2 So were you meaning, expert 1, product attributes in terms of the
14 performance of expert 3 and I?

15 Expert 1 No. I think that if you compared the panels for different products they
16 would have different measures of performance because some product
17 panels are a lot harder than others and I'm not sure if we've covered that.
18 I'm not sure if it comes into group performance. It's affected by the
19 product, not by how the panel was trained

20 Expert 3 Complexity of product?

21 Expert 1 Yeah

22 Expert 2 Yes that's a good description

23 KE Is that a measure?

24 Expert 1 Well it has an effect on the measure of performance

25 KE So product complexity?

026 All Yeah

027 Expert 1 Well should that be in the 'other factors' list then?

028 Expert 2 Yes I think it should actually

029 Expert 1 Yeah put it in the 'other factors' KE, sorry

030 KE So you've taken into account the analysis of performance?

031 Expert 1 It would be a consideration, yeah

032 KE So you can say "they're not getting it but this is a very complex product"

033 Expert 1 Yes that's what we found

034 KE OK, line 25, end of scale values

035 Expert 1 Oh this is in scale development. End of scale words like weak, slight and
036 strong. It's not in the right place

037 KE So it goes back with scale development?

038 Expert 1 Yeah

039 KE Variation, consistency and ANOVA came up. What are you measuring?

040 Expert 2 They're tools to measure the panel's consistency

041 Expert 3 Yes there's reproducibility and then there's consistency

042 KE OK what do you want to call this process? There's the statistical process
043 of measuring variation

044 Expert 2 Yes but we do other things like just graphs on the panel

045 Expert 3 Yeah and there's attribute consistency, time consistency, the individual
046 consistency

047 KE OK so consistency is the word?

048 Expert 1 I'm happy with that

049 KE And you two had consistency anyway

050 Expert 2 Mmm

051 KE OK, F sets?

52 Expert 1 That is F ratio. That's a typographical error on your part!

53 KE I'm sorry. Could you explain F ratio to me please?

54 Expert 1 It's an analysis of variance method. You should know that!

55 KE OK so it measures the consistency thing does it?

56 All Yes

57 KE Fully trained. We had competency come up. This was actually covered in
58 the 'decision' part

59 Expert 3 No that's not a performance measure

60 KE Not going to stand up?

61 All No

62 KE Range?

63 Expert 3 Yes we look at the range of the scores

64 KE Are you all happy with including range as a measure

65 All Yes

66 Expert 3 Test exposures, expert 3, you mentioned that

67 Expert 1 Test exposures!

68 Expert 3 Yes, that's something I do myself. I don't know. I'll have a look

69 Expert 2 Another thing, just while expert 3 is looking, is not only range but the plot.
70 I'm not really sure how to say it but you might plot it out and 5 people got
71 a 2 and 2 got a three you know. It's a box plot

72 KE So do you want to include box plots as a performance measure?

73 Expert 2 Perhaps we could just put graphics or graphical representations or
74 something

75 Expert 1 Yes

76 KE Are you happy with that expert 3?

77 Expert 3 Yes

78 Expert 1 I'm sorry, KE, but I'm going to have to go

79 Expert 2 Did we mention star charts. Like a lot of them don't understand standard
80 deviation so we give them a gold star for good ones and a red star for bas
81 ones, that sort of thing

82 KE How about we include that in the graphical representation side of feedback

83 Expert 1 The only other one that I haven't matched up with the others on is the
84 training environment and I don't know whether you two agree with that or
85 not. It's where you're training some group in a noisy, smelly area compared
86 to a nice quiet one

87 KE So are you happy with that one?

88 All Yeah

89 KE So very quickly, test exposures, just to finish that one...

90 Expert 3 The actual line that you refer to, that word is not there and I don't actually
91 remember what I meant by that. I think what I mean is like half-blind
92 scenarios

93 KE Yep. Now line 50 is population size, line 51, line 52 you all agree on. Shall
94 we leave those in there?

95 All Yeah

96 KE Panel/product attributes, expert 2 and 3 mentioned

97 Expert 3 That's probably product complexity isn't it?

98 Expert 1 Yes

99 KE Is that the term you want to use, complexity?

100 All Yes

101 KE What about the panel attributes like pregnant people and things like that?
102 Is that a separate one?

103 Expert 3 Oh that's a different one

104 KE Is that panel attributes?

105 Expert 2 Yeah, that's like panel make-up

106 KE Panel make-up?

007 Expert 3 Yes, it's the dynamics of the panel really

008 KE OK, happy with that?

009 All Yep

010 KE OK, commercial pressure, is that an 'other consideration'

011 Expert 3 Yes, commercial pressure and time pressure

012 Expert 1 Yes, it's important

013 KE OK, good. How about panel expense? The cost of actually raising and
014 training the panel

015 Expert 2 Yeah, that's an 'other consideration'

016 Expert 3 Yes, agreed expert 1?

017 Expert 1 Yes I agree, but what I'm wondering is why I didn't say it!

018 KE Sensitivity. Over-sensitised, fatigue, that's the sort of context that it came
019 up in

020 Expert 3 Yes, that's part of that time-tabling thing

021 KE Training sessions?

022 Expert 3 Yeah

023 Expert 2 Yes it is when you talk about over-working them

024 KE Right, so training sessions covers that then?

025 Expert 1 Yes

026 KE Compusense. Some of you mentioned this as it came on line during the
027 project. Should we mention it or just drop it?

028 Expert 3 Well I think that we should say data handling...

029 Expert 1 Data gathering, methods of data collection

030 KE OK. That's it. What I'm going to do now is re-draw this chart and I'll drop
031 you back a clean copy and that will complete this stage. What I'm going to
032 do now is go back and start modelling against a library of generic task
033 models that have been developed overseas and I'll probably just drop
034 something in for you to have a look at and make any further comment on.

35 By and large though, most of what I require from you now for this project,
36 as far as I will take it is done and it's just a matter of me putting it
37 together now

38 Expert 1 I'm here for the rest of July then I've got a new job, but I can come in if
39 necessary

40 KE Don't worry. I'll get you!

41 Expert 1 I had a feeling you would

42 **Duration: 1 hour 20 minutes**

Appendix G

Expert Biographical Details

EXPERT 1

Age: 33 years.

Gender: Female.

Education: Master of Horticultural Science (Food Science)
Bachelor of Home Science
Bachelor of Business Studies (Human Resource Management)

Work Experience: NZ Dairy Research Institute
1991 - present TQM Facilitator - Executive
1989 - 1991 Section Manager - Sensory Evaluation
1984 - 1987 Research Technologist

Specific Training: Women in Management - 1988
Project Management - 1989
ASTM E18 Committee - 1987
Advanced Management Programme (NZIM) - 1994

Invited speaker at a range of industry conferences.

EXPERT 2

Age: 29 years.

Gender: Female.

Education: NZ Certificate of Science (Food Science)
Massey Papers in Business communication, business organisation and marketing.

Work Experience: 1983 - present NZ Dairy Research Institute
Sensory Evaluation Section. Research Technologist.

Specific Training: Women in Management - 1993
Consumer Testing, USA - 1994
SAS programming - 1991
SAS graphics - 1992
SAS advanced graphics - 1992
Managing the Nineties - What works for Professional Women
Service Management Workshop

EXPERT 3

Age: 25 years.

Gender: Female.

Education: Master of Food Science (under completion)
Bachelor of Consumer and Applied Science (Food Science and Consumer Food Science)
Massey paper - Applied English

Work Experience: 1993 - present NZ Dairy Research Institute, Sensory Evaluation Section, Research Technologist.

1990 - 1993. Research Assistant.

Specific Training: Women in Management - 1993.
Statistics in Sensory Evaluation, USA - 1994.
Understanding Flavour Quality Conference, UK - 1992.
Experimental design and statistical analysis - 1992.
Survey design - 1992.
SAS programming - 1991.
SAS graphics - 1992.
SAS advanced graphics - 1992.
Quality assurance systems - 1990.