Frontline Learning Research
An official journal of EARLI

Frontline Learning Research Vol. 5 No. 3 Special Issue (2017) 94 - 122
ISSN 2295-3159

Examining expertise using interviews and verbal protocols

Margje W. J. van de Wiel

Maastricht University, the Netherlands

Article received 4 May / revised 2 March / accepted 23 March / available online 14 July

Abstract

To understand expertise and expertise development, interactions between knowledge, cognitive processing and task characteristics must be examined in people at different levels of training, experience, and performance. Interviewing is widely used in the initial exploration of domain expertise. Work and cognitive task analysis chart the knowledge, skills, and strategies experts employ to perform effectively in representative tasks. Interviewing may also shed light on the learning processes involved in acquiring and maintaining expertise and the way experts deal with critical incidents. Interviews may focus on specific tasks, events, scenarios, and examples, but they do not directly tap the representations involved in task performance. Methods that collect verbal protocols during and immediately after task performance better probe the ongoing processes in representing problems and accomplishing tasks. This article provides practical guidelines and examples to help researchers to prepare, conduct, analyse, and report expertise studies using interviews and verbal protocols that are derived from thinking aloud, dialogues or group discussions, free recall, explanation, and retrospective reports. In a multi-method approach, these methods and other techniques need to be combined to fully grasp the nature of expertise. This article shows how the cognitive processes in data collection constrain data quality and highlights how research questions guide the development of coding schemes that enable meaningful interpretation of the rich data obtained. It focuses on professional expertise and provides examples from medicine including visual tasks. This comprehensive review of qualitative research methods aims to contribute to the advancement of expertise.

Keywords: expertise, interviews, verbal protocols, cognitive processing, analysis

Corresponding author information: Margje W. J. van de Wiel, Dep. of Work and Social Psychology, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: m.vandewiel@maastrichtuniversity.nl, Phone: +31-43-3882171, Fax: +31-43-3884211. DOI: http://dx.doi.org/10.14786/flr.v5i3.257

1. Introduction

For this special issue on "Methodologies for studying visual expertise", the present article will discuss the qualitative research methods of interviews and verbal protocols to examine expertise and expertise development. This article aims to guide students, practitioners, and researchers new to the field of expertise research, or these types of qualitative research, when and how to use these methods to answer their research questions. Starting from a theoretical framework of expertise and cognitive processing in task performance, this article provides practical guidelines so that researchers can prepare, conduct, analyse, and report expertise studies using interviews and verbal protocols. The rationale behind the methods, as well as their strengths and weaknesses, are explained to understand how procedures should be designed to maximise the quality of the data. Although these guidelines for research are applicable to all expertise domains, the focus here is on professional expertise. Most examples will be drawn from medical expertise research, as it has a long-standing tradition of using diverse methods of verbal protocols, and includes various areas of visual expertise. This comprehensive overview of qualitative research methods contributes to the literature by showing why and how the methods can best be used to deliver valid, high-quality verbal data when examining expertise. The literature is reviewed from an analytical and practical perspective to connect different research traditions that shed light on the nature and origins of expertise. Expertise research may add to the advancement of any domain, as careful analysis of task characteristics, performance, and underlying knowledge and cognitive processing is at the heart of improving current practices. This review aims to provide researchers and practitioners who want to embark on this endeavour with fundamental insights into theory and methods that help to further develop the level of expertise in their domain of interest.

The article is organised into five further sections. First, expertise is defined in terms of outcomes, underlying knowledge and processes, and their interaction with task and domain characteristics. Second, the steps in preparing expertise research are discussed, starting with the research questions, familiarisation with the domain of expertise and the specific tasks at hand, the selection of experts and other participants, and the main criteria for choosing between interviews or verbal protocols. Third, the interview method is described and placed within the context of research on work and expertise. Fourth, the characteristics of interview and verbal protocol methods are described in light of the cognitive processing involved in task performance and data collection. Moreover, five methods used to gather verbal protocols to reveal expert task performance are discussed and illustrated in more detail. Finally, a conclusion is provided that summarises the main issues to be considered when designing studies that examine expertise using the qualitative research methods of interviews and verbal protocols.

2. Expertise

Two dominant perspectives on expertise can be distinguished in the literature. The expert-performance approach (Ericsson, 1996, 2004, 2015; Ericsson & Smith, 1991) characterises expertise as the capability to demonstrate reproducible superior performance on representative tasks in a specific domain. The highest expertise level is achieved when individuals are able to go beyond mastery and contribute their creative ideas and innovations to the task at hand. Although years of practice and experience are needed to become an expert, skilled performance and experience alone are not enough. Routine behaviour and full automaticity should be counteracted by gaining high-level control of performance that allows further improvements to be made. In the expert-novice research approach (Chi, Glaser, & Farr, 1988; Chi, 2006a), expertise has been characterised by differences in performance and underlying knowledge between groups with increasing levels of experience in a particular domain. Experts have a large and well-developed knowledge-base, that is tuned to the tasks performed and the problems encountered, and allows fast and accurate performance in routine situations. In more complex situations, they can apply their knowledge flexibly when trying to understand the situation and decide upon further actions.

The obvious similarity in these characterisations of expertise is that they both emphasise routine, automatic versus controlled, deliberate performance that is adapted to the task at hand. Both approaches explain how the development of knowledge and skills underlies expert performance (Feltovich, Prietula, & Ericsson, 2006). Simply said, expertise is the result of activating the right knowledge at the right time (Anderson, 1996). Experts have developed rich and coherent knowledge structures that allow immediate access to the relevant knowledge, strategies, skills, and control mechanisms. Domain-specific task performance is mediated by evolving representations of the task and problem they attend to. This enables experts to perform effectively and efficiently, coordinating automatic thoughts and actions with deliberate thinking. Problem representations guide them in selectively focusing on relevant information and features that novices are not aware of. Moreover, they help experts to carefully monitor and adapt their performance in an ongoing process. Figure 1 illustrates how both incoming information and the experts' knowledge in long-term memory continuously interact to determine the content of working memory in task performance. The capacity of working memory is enhanced by retrieval cues that directly access the relevant parts of experts' knowledge in long-term working memory (Ericsson & Kintsch, 1995), enabling them to coordinate thoughts and actions in cognitive processing. The evolving mental representations in task performance reflect the content of working memory. Experts update their knowledge and skills by means of study, practice and experience. They enhance learning from their experiences by seeking feedback and reflecting upon their performance to find weak aspects in processes and outcomes that might be improved. Expertise development is a gradual process in which the knowledge and skills needed to plan, monitor and evaluate performance are refined during practice. This requires the motivation to improve performance and invest effort in deliberate practice (Ericsson, Krampe, & Tesch-Röme, 1993; Ericsson & Pool, 2016).

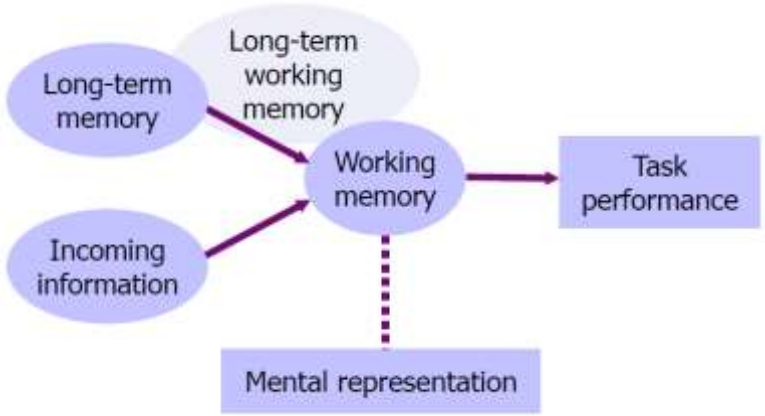

Figure 1. Information processing in task

Although both approaches define expertise in relative terms and focus on tasks representative of the domain, one important difference between them relates to the standards of performance. Whereas the expert-performance approach (Ericsson, 1996, 2015) focuses on top-performance that can be objectively measured, the expert-novice approach (Chi et al., 1988, 2006a) is more pragmatic in comparing novices and students with intermediate levels of training to experienced performers within a particular domain. In professional domains, such as medicine, auditing, law, teaching, software engineering and psychotherapy, it is not as easy to objectively measure performance as it is in domains, such as sports and games, in which clear outcomes (e.g., time, points gained) are available. The experience of the professional and the presence of professional criteria, such as degrees, licenses, memberships of professional organisations, prizes, and teaching experience usually work well to identify experts (Evetts, Mieg, & Felt, 2006; Hoffman, Shadbolt, Burton, & Klein, 1995; Mieg, 2006). There is a notable absence of a 'gold standard' of professional performance that is based on a validated objective outcome measure (Ericsson, 2004; Weiss & Shanteau, 2003; Shanteau, Weiss, Thomas, & Pounds, 2001) and one best solution or approach to a problem may not even exist (Tracey, Wambold, Lichtenberg, & Goodyear, 2014). In medicine, for example, physicians have to make decisions

under conditions of uncertainty when they encounter more complex and rare patient problems. In many other professions, experts are confronted with uncertainty and new situations that require performance on the edge of what they may accomplish based on their knowledge and skills (Klein, 2008; Salas & Klein, 2001). Experts, furthermore, play an important role in the advancement of their domain and in setting (new) standards for performance (Boshuizen & van de Wiel, 2014; Ericsson, 2009, 2015; Evetts et al, 2006; Lesgold, 2000). In summary, how expert performance can best be defined depends on several factors including the domain, the tasks, and the type of problems to be solved.

The accumulated body of knowledge and skills available in a domain constrains the level of performance that can be acquired by individuals. Shanteau (1992) found in his analysis that performance is better in structured domains in which incoming information is static, problems are predictable, conditions are similar, tasks are repetitive, and objective analysis, feedback and decision aids are available. In these structured domains, individuals have more chances to learn and improve their performance as compared to less structured domains, which do not share these task characteristics. Kahneman and Klein (2009) discuss how intuitive expertise, i.e., automatic accurate judgment, can only be developed in high-validity domains in which the environment is predictable via the recognition of a set of cues. If professionals are given adequate opportunity to practice, they can learn the causal structure and/or the statistical regularities that enable this recognition process. Classical examples of ill-structured or low-validity domains include wine tasting, stock broking, and clinical psychology, all domains in which judgment is inconsistent. Ericsson (2014, 2015; Ericsson & Pool, 2016) argues that professional performance can only be improved by searching for and identifying reproducible superior outcomes, which can then be used to guide deliberate practice. Libraries of problem situations with known outcomes and simulators enable intensifying practice with immediate feedback for problems that are uncommon or have high-stakes in real practice. Research on expertise contributes to the development of a domain and the performance levels that can be achieved by professionals on essential tasks.

3. Examining expertise

When examining expertise, first the research question needs to be clearly formulated: “What do you want to know about expertise?” As expertise is based on the development of a well-organised body of knowledge that determines the processes and strategies used in task performance, the most obvious questions are “What knowledge and skills underlie expert performance and how do they develop?”. The research question can focus on the representations of the problem to be solved, or the task to be performed and how these differ between novices and experts (e.g., Chi, Feltovich, & Glaser, 1981; van de Wiel, Boshuizen, & Schmidt, 2000), or change as a result of practice (Boshuizen, van de Wiel, & Schmidt, 2012). But it may also focus on the knowledge and strategies used in problem solving and in performing a task (e.g,, Boshuizen & van de Wiel, 1999; Diemers, van de Wiel, Scherpbier, Baarveld, & Dolmans, 2015; Gilhooly et al., 1997; Lesgold et al., 1988; Kok et al., 2015). It can focus on the learning processes and how teaching and instruction can help novices to become experts (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Kok, de Bruin, Robben, & van Merriënboer, 2013). It may also focus on the activities experts engage in to learn from their experience, further develop and maintain their expertise (e.g., Ericsson et al., 1993; van de Wiel, Van den Bossche, Janssen, & Jossberger, 2011), and the self-regulations skills they apply to plan, control and evaluate their performance. Finally, it may be important to investigate the ways in which the knowledge of experts can fall short by focusing on biases and near-errors and how these might be overcome (e.g., Chi, 2006a; Hashem, Chi, & Friedman, 2003; Elstein & Schwarz, 2002). The general themes addressed by these research questions will require further specification depending on previous research, the domain of expertise, and the research interests.

In the professional domain of medicine, there is a long tradition of expertise research that started with the seminal work of Elstein, Shulman, & Sprafka in 1978. While searching for general problem solving

skills, studies in the early years consistently showed that experts and novices used the same strategy of generating and testing hypotheses in diagnostic problem solving, but that experts generated diagnostic hypotheses faster and more accurately (Elstein et al., 1978; Norman, Eva, Brooks, & Hamstra, 2006; Neufeld, Norman, Feightner, & Barrows, 1981). The accuracy of physicians' diagnoses, however, was found to be case-specific and tied to the domain of clinical experience (Elstein et al., 1978). Research was then directed at uncovering the nature and organisation of knowledge underlying physicians' performance in interaction with the patient cases diagnosed. Results have shown that their large and well-developed knowledge base enables physicians to automatically retrieve the relevant knowledge in routine cases, as well as to analytically process cases that are difficult or evoke a sense of alarm (Elstein & Schwarz, 2002; Stolper et al., 2010). Elucidating the nature and acquisition of medical expertise is still an active research field that contributes to safe patient care and medical education. Studying visual expertise in medicine is a rapidly growing field, as exemplified by this special issue. Imaging techniques are important diagnostic tools that develop quickly and require complex knowledge and skills that need to be learned and assessed (Gegenfurtner, Siewiorek, Lehtinen, & Säljö, 2012). While examining images, bottom-up and top-down processes continuously interact and determine whether significant features are recognised and correctly interpreted. Their knowledge ultimately determines whether experts see what can be detected, and understand what they see. A broad array of research questions is open to investigation in this field.

Having established what you want to know about expertise, the second question that needs to be answered in expertise research is "Who are the experts?" As expertise manifests itself in the context of the tasks that experts engage in, this question is intricately intertwined with the question: "What tasks are critical to the domain?" Depending on the type of performance outcomes available, it may be more or less straightforward to identify experts that consistently show superior performance on representative tasks. A thorough familiarisation with the domain which focuses on the tasks performed is needed to find out what may characterise expert behaviour. If objective outcome measures do not exist, professional criteria that provide social recognition, such as experience, degrees, licenses, job titles, status, and prizes, might be used to define experts (Evetts, Mieg, & Felt, 2006; Hoffman et al., 1995; Mieg, 2006), as well as peer judgments that ask professionals to identify the best performers in their field, or those whom they would go to for advice (Ericsson, 2006a; Kahneman & Klein, 2002; Shanteau, 2002). Other groups of participants must be included to examine in what way experts differ from those who are less experienced within the domain, or those who have worked as long in the profession but are considered to have less expertise. To study the development of expertise, groups with different levels of experience in the domain (ranging from naïve, novice, intermediate, and advanced to expert) are compared to each other. A group of trainees can also be followed on their developmental path towards becoming a professional.

Another crucial question to be addressed in preparing expertise research is "What research method(s) are most suitable to investigate the research questions in this field?". In this article, the focus is on qualitative research methods of interviews and verbal protocols as they play a crucial role in uncovering the characteristics and origins of expertise in any domain. Interviewing is a very straightforward way to initially explore a specific domain of expertise. It can be used to gather information on the relevant tasks undertaken, the knowledge and skills needed to perform these tasks and solve problems, the learning processes involved in education and continuous development, and the pitfalls associated with expertise that need to be dealt with. Interviews deliver verbal protocols as data resulting from the answers to the interview questions. However, to better capture the cognitive processing of experts in task performance, verbal protocols must be gathered that are directly related to the task-specific processes. Expertise researchers have, therefore, developed methods that study, in addition to behavior and outcomes of representative task performance, the thinking processes involved. They do so by probing the experts' underlying representations, knowledge, and reasoning (Chi, 2006b, 1997; Ericsson & Simon, 1980, 1993; Feltovich et al. 2006; Hoffman et al., 1995). These methods provide insight into the content of working memory during, or immediately after, task performance (see Figure 1). Two common methods used to assess online thinking by gathering verbal data during task performance are thinking aloud and discourse analysis of dialogues and group discussions. Three common methods used to gather verbal protocols after task processing are free recall protocols, explanation

protocols, and retrospective reports. Table 1 provides an overview of the qualitative research methods discussed in the present article, and how they are related to both the domain and the task when examining expertise.

Table 1

Overview of qualitative methods used to examine expertise in relation to the domain and the task

Exploration of domain expertise	**Cognitive processing related to a specific task**	
	During task performance	**After task performance**
Interviews	Thinking aloud	Free recall
Focus groups	Dialogues and group discussions	Explanation
		Retrospective reports

To guarantee the quality of the data obtained by interviews and verbal protocols in expertise research, careful preparation is required. The most critical steps in preparing studies using these qualitative research methods are summarised in Table 2. In addition to the steps outlined above, how the protocols will be coded, and how the study will be communicated to the participants, are two important considerations for both interviews and verbal protocols. In relation to interviews, the emphasis is on preparing the interaction with the interviewee, whereas for verbal protocols the emphasis is on selecting and designing tasks that may differentiate expert and novice behaviour. These tasks should reflect the same goal-directed processing as required in real-world tasks. In the following sections, the methods are discussed to provide practical guidelines for developing, conducting, and analysing expertise studies that deliver valid, high-quality verbal data, which can then be reported in a transparent way. In addition, the strengths and weaknesses of these particular methods are highlighted and compared, and specific issues related to expertise research are explained and illustrated.

Table 2

Steps in preparing studies using interviews and verbal protocols to examine expertise

Formulating the research question	
Familiarisation with the domain and tasks	
Determining the experts and participants	
Interviews	*Verbal protocols*
Developing the interview guide	Selecting and analysing the task
Preparing for the role of interviewer	Developing the task materials and instructions
Developing coding schemes for analysing the verbal protocols	
Framing the communication to participants	

4. Interviews

Interviewing is one of the most common methods used to gather information about a given topic. It is a very natural process of inquiry that is used in everyday communication; just think about how often we engage in asking questions and receiving answers. The key to all good interviews is to clearly ask what you want to know, and to make sure that you receive the answer that allows you to know what you want to know. This sentence describes interviewing in a nutshell, highlighting the importance of the research questions, the interview guide, and the role of the interviewer in asking questions and evaluating answers. In relation to expertise research, it is important that this characterisation of interviewing also shows a realistic perspective on research. It assumes that we can come to understand a topic by obtaining relevant information in an objective way by interviewing a representative sample of participants (Emans, 2004; King & Horrocks, 2010). The interviewer wants to reveal what interviewees know, do, think, feel, believe, intend to do, want, or need, and assumes that the interviewee can communicate this during the interview.

The basic processes involved in interviewing for research purposes are well explained by Emans (2004) and summarised in Figure 2. The goal is to reveal the cognitions of the interviewee, as related to a certain topic, i.e., the interviewee's mental processes and the products of these processes, usually in the form of information, knowledge, thoughts, feelings and ideas about the topic. The task of the interviewer is to create a situation and ask questions that motivate an interviewee to connect to these cognitions and verbalise them in a reliable manner. The interviewer must carefully listen to the answers provided, and check whether these answers include the information necessary to answer the research questions. If not, further questions need to be asked. To obtain data for subsequent analysis, the whole interview needs to be recorded and transcribed verbatim, thus without any interpretation of the data.

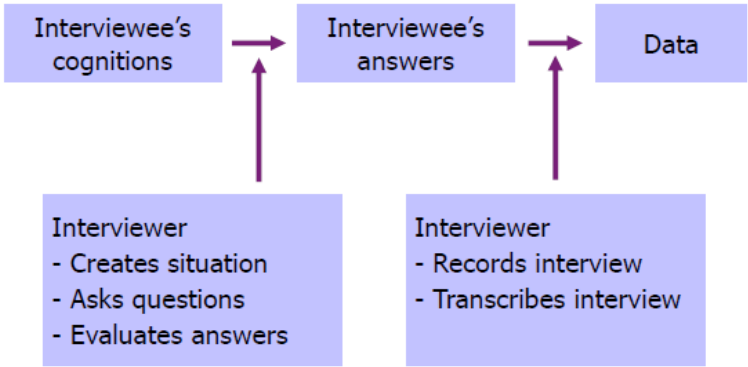

Figure 2. Processes in interviewing (adapted from Emans, 2004).

In the context of work, interviewing is the most common method used to interact with experienced practitioners as subject-matter experts to gather information about all kind of aspects of the work they are engaged in and the vocabulary they use. In job analysis, cognitive task analysis, and knowledge elicitation, interviews are used to yield primary insights into the tasks experts perform, the knowledge and skills underlying their performance, and the conditions that shape their performance. Job analysis focuses on work activities, worker attributes, and/or work context and is used to inform human resource management practices, such as personnel selection, training, and performance management (Bartram, 2008; Sanchez & Levine, 2012). Job analysis is also a first step in job (re)design, workplace and equipment design and organising team work and provides an overview to determine what tasks need to be further scrutinised by task analysis or cognitive task analysis (Chipman, Schraagen & Shalin, 2000; Dubois & Shalin, 2000). Detailed analysis of tasks in terms of goals, actions, and thought processes helps to articulate what is not directly observable but which may be expressed by experts when they are sufficiently guided. In knowledge engineering, knowledge elicitation techniques are specifically designed for this purpose in order to develop expert systems and knowledge management systems (Hoffman et al., 1995; Hoffman & Lintern, 2006;

Shadbolt & Smart, 2015). Interviews in different forms have been applied in a wide variety of professional domains to elicit knowledge from experts. The unstructured interview is often employed in an initial exploratory phase in which investigators familiarise themselves with the domain in an informal setting. In later phases, more structured interviews can scaffold the knowledge elicitation process by focusing on specific events, such as critical incidents in the *critical incident technique* (Flanagan, 1954) and critical decisions made in unusual and challenging cases in the *critical decision method* (Hoffman & Lintern, 2006; Shadbolt & Smart, 2015). Experts can also be asked to respond to an evolving scenario or to specific probe questions in order to systematically unravel their task representations (Shadbolt & Smart, 2015). The literature on job analysis, cognitive task analysis, and knowledge elicitation also emphasises the need for research methods to be combined to achieve a thorough understanding of the job and the tasks under study. As complex tasks often require teamwork, interviewing teams in addition to individual experts can provide further insight into how experts from different disciplines work together, build a shared understanding of tasks and situations, and coordinate and distribute tasks amongst themselves (Salas, Rosen, Burke, Goodwin, & Fiore, 2006). The interview method provides relatively quick access to information from several teams within a domain, as compared to observation in the field or simulation of task performance. Furthermore, as jobs, tasks, equipment, and work roles are not stable but rather continuously developing, experts in the field par excellence may provide valuable perspectives on future developments, and innovations, and insights into novel problems and how to deal with them (Boshuizen & van de Wiel, 2014; Lesgold, 2000; Sanchez & Levine, 2012). In fact, experts shape the advancement of their field, and may share their ideas about these developments, and how to support them by individual and organisational learning, in future-oriented interviews (Bartram, 2008; Evetts et al, 2006; Lesgold, 2000).

Whereas less structured, informal interviews are very suitable for an initial exploration of a domain or topic, semi-structured interviews are most useful in expertise research as they result in more objective data that also allow for comparisons to be made between different expertise groups. These interviews provide enough guidance to structure the conversation, but also enable the acquisition of meaningful information, as the interviewer may interact with the interviewee after an initial open question (Emans, 2004). Focus groups that investigate the opinions and experiences of people while they are interacting in groups are valuable tools to examine expertise for both exploratory and comparative purposes. In focus groups, initial questions guide the interview by inviting participants to share their views. In preparing interviews and focus groups for expertise research, the steps outlined in Table 2 must be kept in mind. In the following sections, the development of the interview guide, the preparation for the tasks of interviewer, and the analysis of the data are described and illustrated. Although face-to-face interviews are most commonly used, and taken as a starting point in the descriptions, the guidelines are largely applicable to interviews conducted in groups and by telephone or Skype (Deakin & Wakefield, 2014; Emans, 2004). In a subsequent section, the focus group method is explained in more detail.

4.1 The interview guide

An interview guide is a script that helps to ensure that the interview is conducted in a standardised way. It consists of an introduction to the study, the body of the interview outlining the main questions and possible follow-up questions, the transitions between questions, and a conclusion (Emans, 2004; King & Horrocks, 2010; Skopec, 1986). In the recruitment phase, participants already receive information about the study that may influence their contributions, and thus, the quality of the data gathered. Therefore, it is good practice to compose this information as the first step of the interview guide. For researchers new to the field of the interview method Table 3 provides an overview of the topics advised to be addressed in the interview guide.

Table 3

General outline of the interview guide

Information sheet
Provide context and purpose of the research; invite the participant; explain the procedure and expectations; explain the costs and benefits for participants; emphasise that participation is voluntary and may be withdrawn at any time; explain how data will be used; ask if there are any remaining questions; provide details of the research team and contact information (check compliance to ethical standards)

Introduction of the interview
Introduce yourself as interviewer; explain the purpose of research; show appreciation for participation; give the information letter and ask for informed consent; give a brief outline of the interview topics and indicate the time involved; clarify the roles and expectations of both interviewer and interviewee; ask permission to record the interview; ask if there are any remaining questions

Body of the interview
Introduction Theme 1
Question 1 + (probes) + follow-up questions
Question 2 + (probes) + follow-up questions
Question 3 + (probes) + follow-up questions
Et cetera
Introduction Theme 2
Question 1 + (probes) + follow-up questions
Question 2 + (probes) + follow-up questions
Et cetera

Conclusion of the interview
Indicate the end of the interview; summarise the main points; thank the interviewee for the valuable contribution; reiterate how the data will be used; provide a debriefing of the study; offer to provide a summary of the results after data processing has taken place; ask if there are any questions or comments

Developing the questions for the interview is an iterative process that is based on the research goals and the specific research questions (Emans, 2004). As described above, the goal is to have a clear idea of what you want to know, and then to ask questions that elicit this information from the interviewees. In interview studies, it is also important to make the need for information explicit by defining the variables to be examined. The choice of variables should reflect a thorough understanding of the domain of study and, whenever possible, should be grounded in theory and based on previous research. Formulating the possible outcomes of these variables assists in the phrasing of the interview questions and the analysis of the data. In fact, the possible outcomes are the type of answers you would expect to receive from the participants in the interview and, thus, provide clear guidelines for formulating the interview questions. An example of a study on the development of medical expertise in professional practice may illustrate this approach. In this study, our exploratory research questions were: "How do physicians learn in, from, and for their daily work, and how deliberate is this learning process?" (van de Wiel et al., 2011). In addition, we wanted to examine differences in workplace learning between three groups of physicians. If we had asked our research questions directly, we would have cued the participants' connotations about learning. This may have led to the physicians limiting their answers based on their interpretation of the concept, and focusing their attention too much on deliberate processes. Based on research on workplace learning, deliberate practice, and self-regulated learning, we identified several relevant work-related activities from which physicians could learn: *problem solving, consultation of colleagues, having differences of opinion, explaining to others, seeking and receiving feedback, evaluating performance, professional development activities*, and *participation in research*. These work-related activities allowed us to formulate more specific research questions that helped us to decide what variables to focus on in the interviews. Table 4 shows how specific research questions,

variables, possible outcomes, and interview questions are related. For example, the specific research question formulated for the work-related activity of *problem solving* concerns what physicians can learn from the problems they encounter. These problems provide a chance to learn because they need to be solved, or dealt with as part of the job. The two most important variables to examine, therefore, are the types of problems encountered (i.e., what topics can physicians learn about?) and how they solve these problems (i.e., in what way do they learn?). Some answers can already be anticipated and listed as possible outcomes in order to guide the formulation of the interview questions on this topic. In our study, an additional question asked for an example of a specific problem and how it was solved in order to illustrate and corroborate the previous answers. A subsequent research question that we addressed was what physicians can learn from solving problems by consulting colleagues. The variables of interest follow-up on the problem solving strategies previously mentioned by the interviewees. As we expected the consultation of colleagues to be an important strategy, we first allowed participants to mention it spontaneously, and then elaborated on the topic in the next interview question. In this study, we started with general questions and moved on to more specific learning activities. Moreover, we were careful to introduce the study using the term professional development and not learning, as interview questions prime the way in which participants think and answer.

Table 4

Relationships between specific research questions, variables, possible outcomes, and interview questions. Examples are taken from a study on physicians' learning in the workplace (van de Wiel, et al., 2011)

Specific research questions	Variables	Possible outcomes	Interview questions
What can physicians learn from the problems encountered in their practice?	Types of problems encountered	- Problems in diagnosing patients - Problems in deciding on the treatment - Problems in interacting with patients	1. What problems do you usually encounter when diagnosing and treating patients?
	Types of strategies in problem solving	- Asking colleagues for advice - Searching the literature	1a. What strategies do you use to solve these problems? *(if needed can you explain why?)*
		- Example 1 - Example 2 - Et cetera	1b. Can you give an example of a problem and how you solved it?
What can physicians learn from consulting colleagues?	Kind of situations in which advice is asked	- Lack of knowledge - Being uncertain	2. In what situations do you choose to ask a colleague for advice?
	Frequency of advice seeking	- Daily - Weekly - Monthly	2a. Can you indicate how often you ask for advice? *(ask to specify)*
	Reasons for advice seeking	- Direct need for patient care - Learning	2b. Why do you ask for advice? *(probe for purpose/goal)*
	Et cetera		

The type of variables tapped into by the interview questions constrains the answers that can be expected. With closed questions, answers are restricted to a closed set, such as years of experience, age, or

frequency (e.g., question 2a in Table 4). In bipolar questions, this set is restricted to the answers yes or no. These questions are mostly asked in order to get a clear answer that can be followed up. In open questions, the interviewees are invited to share their perspective on a topic (e.g., questions 1, 1a, 1b, 2, and 2b in Table 4). These questions must be unambiguously formulated, ask about one specific topic at a time, and should not be leading, i.e., should not suggest a particular answer option. Interviewers must be neutral and not introduce bias by showing their own opinions, ideas, or feelings, or by suggesting answer options by providing examples or disclosing expectations. The goal is to obtain valid objective data and these types of suggestive questions may influence the thought processes of the interviewee. In answering open questions, interviewees must be encouraged to bring forward what is most prominent in their mind from their own frame of reference. The research team must critically review the interview guide and pilot the interviews with representatives of the target group. This will optimise the informational value of the data collected and prevent the use of restrictive and leading questions.

4.2 The tasks of the interviewer

Interviewing is a complex task in which the interviewer has to obtain the required information as efficiently as possible. The interviewer must create an interview situation in which the interviewee feels encouraged to speak up, but at the same time, the interviewer must remain in control by asking questions, evaluating the answers, and probing for more elaborate or meaningful answers if necessary (Emans, 2004; Skopec, 1986). This means that the interviewer must build up a positive relationship with the interviewee to make him or her feel at ease, and be well-prepared to guide the interviewee through the questions in an unobtrusive manner. While it is vital to standardise the situation over different interviews to obtain objective and comparable data, a natural conversation will usually occur if the interviewer is clear, kind, and genuinely interested in a professional way. The interviewer observes the interviewee and listens carefully to the answers by keeping the final goal in mind: gathering valid, complete, relevant and clear answers that reflect the interviewee's cognitions and provide the information needed for the study. The interaction with the interviewee is steered by verbal and non-verbal probing. Nondirective probes serve to fuel the conversation by encouraging interviewees to continue without interrupting their line of thought. Effective nondirective probes include: silence accompanied by non-verbal signs of attention; neutral phrases, such as "um hmm", "oh", "yes", "interesting"; rephrasing part of the answer; reflection of feelings that cannot be neglected by showing understanding; making brief summaries of what has been said to check the main points; and general elaborations such as, "Can you tell me more?" and, "Could you explain further?". Direct probes more actively intervene in the flow of the conversation and are used to focus attention on specific topics. Common directive probes include: elaborations that ask for specifications; clarification questions when answers given are imprecise or not well understood; repetition of a question when it appears that the interviewee has not understood the question or avoids answering it; and confrontation when answers seem inconsistent. The interviewer must take the lead and coordinate questioning and probing with appropriate non-verbal behaviour in terms of eye-contact, facial expressions, gestures, and position, to obtain the information required and maintain a positive atmosphere throughout the interview.

4.3 Analysing and reporting the data

Analysing interviews can be relatively straightforward, if during the translation of the specific research questions into variables and interview questions in the preparation phase (see Table 4), the anticipated answers were matched to the intended results (Emans, 2004). A good match guarantees the internal validity of the study (Neuendorf, 2002). The analysis starts with transcribing the interviews verbatim and proceeds by completing the list of possible answers anticipated while preparing the interview with the answers actually given by the interviewees. The more open the questions, the more diverse the answers can be. The more limited the possible answer set, the easier it is to create an overview of the results. For closed questions asking for numerical data, such as work experience, age, and frequency, the results can be

processed as they are in quantitative studies, and groups can be compared. For the open questions, researchers need to categorise the answers per variable using content, thematic, or template analysis (Neuendorf, 2002; King & Horrocks, 2010; Brooks, McCluskey, Turley, & King, 2015). Categorisation starts by developing a coding scheme indicating all emergent answer categories per variable. Coding is usually an ongoing and iterative process. A team of researchers reads and codes the transcripts and discusses these codes as new themes and subthemes in the answers emerge and are agreed upon. It is vital to code all relevant parts of participants' answers. Irrelevant answers, i.e., those that do not relate to the research questions, may be categorised under a separate code. This helps the researchers to check those answers and be open-minded to the possibility of finding unexpected emergent themes in the analysis, related to questions answered throughout the interview. The coding can be done manually or can be facilitated by software programs for qualitative analysis. After the coding of all transcripts is complete, the frequencies per code can be listed to indicate the most common answers to each question, as well as the exceptional ones. A good overview shows the main themes and subthemes and helps to uncover patterns in the data as well as any differences between participant groups. In essence, the analysis and reporting of interview data is a process of data reduction and summarisation. Illustrative quotes from participants' answers enrich reporting by giving insight in how participants typically expressed themselves. The description so far may be very abstract, but an example will show that the approach is actually very pragmatic.

In the study we conducted about physicians' workplace learning (van de Wiel et al., 2011), we started by categorising the answers per interview question and later grouped them per variable and per specific research question (see Table 4) in order to meaningfully report the data. For example, regarding the interview question about what problems participants encountered when diagnosing and treating patients (Question 1 in Table 4), we found that most participants encountered problems that could be categorised as problems with diagnosis, choosing diagnostic tools and treatments, interaction with patients, and practical organisational issues. Participants gave specific examples of some problems, and these were summarised. The ways in which they solved these problems were also categorised and summarised. The answers to these questions were often intermingled, and just as the interviewer had to make sure that all topics were addressed, the researchers had to combine all answers in the analysis. As theory and previous research provided the basis for our specific research questions and variables, we chose to report the data by presenting these as themes and subthemes in the results section. Our intention was to summarise what had been said by the participants and to indicate to what extent they concurred and differed on themes and subthemes. In our reporting of results, we also referred to characteristic quotes. Analysis was an iterative process in which two coders consecutively categorised sets of data. These categories were reviewed and critically discussed by the research team.

In a follow-up study we wanted to quantify to what extent the physicians were deliberately engaged in the work-related learning activities and relate this to other variables (van de Wiel, & Van den Bossche, 2013). In a second step, we therefore analysed the data used in the van de Wiel et al. (2011) study from this perspective. The analysis approach taken in this study illustrates how rich verbal interview data are, and how these data can be analysed in different ways depending on the specific research questions. The themes reported in the results of the qualitative analysis of the first study (i.e., work-related learning activities in medical practice) were used as variables to be coded to explore the extent of deliberate engagement in these activities in the second study. In an iterative process, three researchers coded a subset of the interviews until they could reliably distinguish three levels of deliberate practice for each variable: (0) not engaged in learning activity, (1) engaged in learning activities inherent to the job, such as solving a problem, and (2) engaged in deliberate practice as indicated by showing greater motivation and effort for learning to improve competence. Clear definitions of codes guided the continuation of coding by one researcher, who consulted the others when in doubt. Two themes that pervaded the entire interview were added as variables: *reflection on diagnosis and treatment* and *planning learning activities*. The categorisations were reported in a table to allow comparison between the medical residents and the experienced physicians participating in the study. The table displayed the frequencies of both groups' learning activities, representing the ten variables at each of the three levels of deliberate practice. The outcomes were described in the text and illustrated by quotes.

4.4 Focus groups

Focus groups are the most common method used to investigate the opinions and experiences of carefully selected groups of people with regard to all kind of topics and across a wide range of fields (Morgan, 1996; Krueger & Casey, 2015; Stalmeijer, McNaughton, & van Mook, 2014). In relation to expertise research, this method is a valuable tool that can be used to gain insight into how different people experience a task, situation, or phenomenon. It can be used both to explore a topic and to compare different groups. Only a few initial questions are needed to open up the discussion and encourage participants to share their view on the topic at hand. The advantage of this group interview method is that as people interact, they discuss and analyse the topic from different perspectives, ask each other questions, and may refine their views. A moderator guides the discussion and, as in individual interviews, is responsible for making the participants feel at ease while at the same time probing them to specify their contributions in order to obtain relevant data. A co-moderator usually assists in this process. The discussions are transcribed and then analysed in a bottom-up way by identifying themes and subthemes throughout the text that are relevant to the specific research questions. The researchers also look for relationships between these themes. At least two coders analyse the data independently and then critically review the coding scheme until they reach agreement. The researchers write a summary of the findings that may then be sent to participants to check whether they have suggestions for adjustments that would better represent their discussion. The summary also pinpoints issues that need further clarification and can be brought up at the next focus group. Usually 3-4 rounds of focus group discussions with 5-8 participants in the target group are needed. Extra groups are added as long as new information emerges in the discussions, i.e., until data saturation has been achieved. A synthesis of all themes and subthemes coded in the different groups is the basis for reporting the data. Quotes can illustrate characteristic utterances.

The focus group method has been frequently applied in medicine, often for purposes of medical education (Stalmeijer et al., 2014), for example to gain insight into how to improve study arrangements (e.g., de Leng, Dolmans, van de Wiel, Muijtjens, & van der Vleuten, 2007; van de Wiel, Schaper, Scherpbier, van der Vleuten, & Boshuizen, 1999), and how the transition from medical school to clinical practice is perceived and can be supported (Prince, van de Wiel, van der Vleuten, Boshuizen, & Scherpbier, 2004). In medical expertise research, this method has been used to understand how general practitioners approach the diagnostic task and what role non-analytical reasoning plays in their diagnostic process (Stolper et al., 2009).

5. Verbal protocols

Verbal protocols add another dimension to the examination of expertise as they deliver verbal data in relation to cognitive processing either during or directly after task performance. Interviews are limited in that they provide self-reports by eliciting cognitions about task performance and expert behaviour in a general way. This may induce participants to interpret their own processes from an evaluative perspective and lead to reconstruction and generalisation of their memories of specific task performance (Ericsson & Simon, 1980; 1993; van Someren, Barnard, & Sandberg, 1994). To minimise these effects, verbal protocol methods aim to capture the processes in performing representative domain tasks by tapping the content of working memory during or immediately after task processing (see Figure 1). This diminishes participants' opportunity to theorise and rationalise what they do and keeps the time delay in processing and verbalising to a minimum. However, requesting participants to verbalise their thoughts while they engage in a task may interfere with natural processing. Interference may also occur when participants know beforehand that they will be asked to report back on their task performance. It is necessary, therefore, that studies using verbal protocol methods are carefully designed to capture the natural task processes and problem representations. The formulation of task instructions and selection of tasks and problem situations are critical to ensure that expert behaviour can be demonstrated, and contrasted to novice behavior, in goal-directed, realistic task performance. Table 5

shows how the different methods discussed in this article score on the most important criteria in determining whether verbalisation of cognitions impacts the validity of the verbal data gathered.

Table 5

Advantages and disadvantages of the qualitative methods of interviews and verbal protocols in relation to task processing criteria that may impact the validity of verbal data

Method	Specific task performance	Interference with task processing	Inducing interpretation	Reflection of working memory
Interviews	-	-	+/-	-
Focus groups	-	-	+/-	-
During task				
Think aloud	+	+/-	-	+
Dialogues and group discussions	+	-	-	+
Probing	+	+	+/-	+/-
After task				
Free recall	+	-	-	+
Explanation	+	-	-	+
Retrospective reports	+	-	-	+
Probing	+	-	+/-	+/-

Two methods that lie at the intersection of interviews and verbal protocols probe participants' cognitive processing by means of questions during or after specific task performance. As depicted in Table 5, one advantage of these probing methods is that cognitions can be examined in direct relation to the task at hand, yielding more specific and precise information than interviews. The disadvantage of probing during task performance is that questioning interrupts participants' thoughts and actions, altering their normal task processing. This may even encourage participants to adopt an interpretative mind-set (Ericsson & Simon, 1980, 1993; van Someren et al, 1994). Probing after task performance overlaps with interviews that focus on specific tasks, events, scenarios, and examples as used in knowledge elicitation techniques (Hoffman et al., 1995; Hoffman & Lintern, 2006; Shadbolt & Smart, 2015), as well as with some types of explanation protocols and retrospective reports. This method can deliver very valuable information regarding the research questions. As discussed in the section on interviews, this is dependent upon the way the questions are phrased and embedded within the interview guide. In expertise research, data gathered by interviews and verbal protocols may complement each other as participants' overall cognitions about their expertise domain can be combined with assessments of task-specific processing and outcomes.

As expertise is domain and task specific, the selection of participants, tasks, and particular problems to solve are crucial steps in setting up verbal protocol studies that examine expertise (see Table 2). Both task characteristics and the experts' knowledge and experience determine the cognitive processes and evolving representations in task performance (see Figure 1). If a representative domain task has been chosen, e.g., diagnosis in medicine, the next step is to decide upon the problems to be solved and the presentation format. In medicine, patient cases that reflect a consultation with a physician can, for example, be summarised in a brief description and might be supplemented with information from the patients' record to mimic the situation in real practice. Most important here is that the experimental situation captures the essentials of the task and the problem under investigation. In order to identify characteristics of expert behaviour, cognitive processing, and outcomes, as well as differences with other groups of participants, problems may be presented at various levels of difficulty. Task materials and conditions may also be manipulated to

investigate the effects of changing normal processing. In routine problems, experts are expected to automatically activate the right knowledge, but in more difficult problems and under special conditions, they will coordinate automatic thoughts with analytical thinking. The amount of problem information presented and the time in which the information becomes available also impact cognitive processing: the more information and time involved, the more coordination is required, and the more elaborate and deliberate thinking will be. Factors such as these that are related to the nature of the task performed are also emphasised in the cognitive continuum theory (Custers, 2013; Hamm, 1988; Hammond, Hamm, Grassia, & Pearson, 1987). This theory situates most thinking somewhere in between intuition and analysis, which are conceptualised as two ends of a continuum of cognitive processing. Where on the continuum thinking falls, depends on specific task characteristics. This theory provides a framework for analysing tasks in relation to processing requirements. Verbalisation theory, in addition, provides a framework for analysing in what way verbalisation of the content in working memory influences task processing (Ericsson & Simon, 1980, 1993). If information and knowledge are already represented in a verbal format, they only have to be vocalised. If, however, they are represented in a visual or motor modality, encoding of the representations into verbal format is required. This demands extra cognitive processing that may alter the way in which the task normally proceeds. In preparing verbal protocol studies, both the task and problem characteristics need to be well thought out, and piloted to ascertain that the research questions can be answered. These characteristics may also be manipulated to test specific hypotheses. When combining verbal protocol methods in one study, researchers need to be careful that instructions given and procedures followed do not influence subsequent task processing.

In conducting verbal protocol studies, task performance must be monitored and verbalisations need to be recorded and transcribed verbatim (Chi, 1997; Ericsson & Simon, 1980, 1993; van Someren et al, 1994). The outcome measures of the task comprise the dependent variables that are to be used to assess differences in expertise between groups or improvement over time. These may be complemented with quantitative processing measures, such as time used to solve or to explain a problem. The verbal protocols provide rich data on the underlying cognitions in task performance that need to be coded and interpreted to answer the research questions. Analysing the data from a clear perspective enhances the acquisition of valuable, objective information and may reduce the workload. A coding scheme depicting the variables (e.g., knowledge used) and the related coding categories (e.g., biomedical and clinical knowledge), including definitions and examples of utterances per category, needs to be developed to guide data analysis (Chi, 1997; van Someren et al., 1994). The coding scheme can be based on theory, previous research, and/or cognitive task analysis or can be developed in a bottom up way (Chi, 1997; Hsieh & Shannon, 2005; van Someren et al., 1994). An important issue to decide upon when developing the coding scheme is the unit of analysis used, as this may vary from a word, single unit of information or proposition, to a clause, reasoning chain, or turn in a discussion, and is to be guided by the research goals. Particularly in theory-driven research, it is good practice to develop the coding scheme using a set of pilot protocols and to test the hypotheses on another sample of protocols. In more exploratory research, the development of coding schemes may lead to theory-building and hypotheses generation. During analysis, the audio- or video-recordings can be listened to or watched, if it is necessary to improve understanding and interpretation. The coding of verbal protocols is an iterative process, in which coders must seek agreement in order to obtain reliable data. Although the analysis of verbal protocols is qualitatively in nature, the data may be quantitatively described by tallying the number of utterances per coding category (Chi, 1997; Krippendorf, 2012; Neuendorf, 2002). Such quantitative descriptions help to create an overview, reduce subjectivity in interpretation, and find and report meaningful patterns in the data, while protocol fragments show examples of the coded categories in relation to each variable.

Recapitulating the steps to be taken in preparing expertise research using qualitative methods (see Table 2) clearly shows how important it is that the study design is guided by the research questions and by the strengths and weaknesses of the methods employed. Understanding the ways in which various methods affect cognitive task processing in data collection helps to establish what interview or verbal protocol method to use. In designing verbal protocol studies, the next critical step is to choose task characteristics and

requirements in a manner that can reveal expert behaviour under conditions reflecting or manipulating the essence of the task. This step is closely tied to the development of a coding scheme in which it is anticipated how each variable of interest can be measured in the verbal protocols. The final step of communicating the study to participants must comply with general research guidelines, as described in the outline of the interview guide (see Table 3). In the following sections, the five verbal protocol methods of thinking aloud, dialogues or group discussions, free recall, explanation, and retrospective reports will be discussed. Drawing from research on medical and visual expertise, examples will be provided of specific aspects of the design and analysis of each method.

5.1 Think-aloud protocols

The think-aloud method has been frequently used to investigate the knowledge and processes in task performance and is well-documented in the literature (Chi, 1997; Ericsson, 2006b; Ericsson & Simon, 1980, 1993; Hassebrock & Prietula, 1992; van Someren, et al., 1994; Shadbolt & Smart, 2015). Participants are asked to say everything that comes to their mind while they engage in a task. The rationale behind this method is that the verbalised thoughts reflect the evolving mental representations in working memory during task performance (see Figure 1). Verbalisation of thoughts does not change the sequence of actions, and usually does not disturb the cognitive processes engaged in, but merely slows down these processes. In some tasks, however, verbal encoding and vocalisation of information interferes with natural task performance. This is because the increased load on working memory makes it difficult to keep up with the flow of information that needs to be attended to, i.e., cognitive processing cannot be slowed down to successfully accomplish the task. Verbalising is easiest when thoughts are already verbally represented, and requires extra cognitive effort in visual and motor tasks. In highly skilled and expert performance, cognitive processes are largely automated and think-aloud protocols will only reveal those thoughts that consciously come to mind. This method, then, shows what knowledge is activated, which parts of cognitive processing are automated, and when deliberate, analytical thinking is involved in specific groups, tasks and problem situations. Natural thinking in cognitive tasks can be easily disturbed by task instructions. It is therefore important that these instructions are formulated in such a way that participants are not tempted to explain or justify what they do to the experimenter. Practicing the think-aloud procedure with participants maximises the chance of obtaining valid data. When verbal materials are presented, these need to be read out loud to facilitate expressions of thoughts during information intake and signal the cues that trigger these thoughts. The data can be systematically analysed and reported in many different ways, depending on the research goals.

In medicine, the think-aloud procedure has, for the most part, been applied in diagnostic tasks to reveal the knowledge and reasoning involved. Hassebrock & Prietula (1992) made a detailed analysis of diagnostic reasoning, focusing on *knowledge states*, *conceptual operations*, and *lines of reasoning* to explore different types of diagnosis, the cognitive activities engaged in (e.g., *data examination*, *data explanation*, *hypothesis evaluation*, *meta-reasoning*), and the links between patient cues, pathophysiological conditions and hypotheses, respectively. The coding scheme they used was very elaborate allowing for precise statements on the knowledge representations and the (causal) lines of reasoning in diagnosing cases of congenital heart disease. These statements could be compared with expert models of reasoning on the topic. It is a good illustration of how cognitive task analysis may guide the interpretation of qualitative data. However, although it is tempting to carry out this type of detailed analysis when data are so rich, it might be more practical to focus on some of the main inferences made, as in a study conducted by Gilhooly, et al. (1997). They asked participants to diagnose eight ECG traces while thinking aloud. A computer program first listed all of the technical terms used by participants and an expert categorised these words into three coding categories. Subsequently, the program counted the number of words indicating *trace characteristics*, *clinical inferences* and *biomedical inferences*. This method enabled the researchers to compare the knowledge used in visual diagnosis between different expertise groups and across ECG traces varying in difficulty, as well as between the think-aloud protocols and explanation protocols that were collected a week later. In this special issue, Helle (2017) discusses the relationship between eye tracking data and verbal data.

5.2 Discourse analysis of dialogues and group discussions

Recording the natural discourses of collaborating experts or of students and teachers in education is a valuable method that can be used to gain insight into online group decision making, problem solving, and learning (Chi, 1997; Salas et al., 2006). It shows what issues participants attend to, how they regulate their discussions and task processes, how they interact, what kinds of knowledge and strategies they use, what they might learn, and what they can improve. The main task of researchers is to select a representative sample of meetings in which participants engage in knowledge sharing as part of their work or training, audio- or video-record these meetings, transcribe them verbatim, and then analyse what has been said. Participants might notice that they are being observed in the very beginning, but as soon as they start their tasks, they will proceed as usual. This method delivers very rich data, and reducing the data to manageable proportions in coding is a challenge, and a process that will be guided by the research goals. Analysing and describing the data at different levels of detail, for example, in terms of the sequence of actions taken in discussing a patient, the type of patient problems discussed, and the content of a sample of these discussions, helps to create overview and pinpoint the issues of interest.

Patel and colleagues investigated team problem solving and decision making in the complex environment of hospital intensive care units. In their research, work domain analysis and communication patterns between attending physicians, residents, nurses, and consulted specialists provided an invaluable framework that could be used to analyse the content of the individual contributions and understand the processes in context (Patel & Arocha, 2001; Patel, Kaufman, & Magder, 1996). The researchers particularly focused on discussions that took place during morning rounds. During these rounds the team discusses the patients on the ward, evaluating each patient in detail and planning future actions. They enriched their analysis by examining complementary data from patient charts, recordings of morning lectures, and interviews with participants. They segmented the protocols into episodes that distinguished between subsequent phases in the discussions, and further segmented these phases into thematic idea units. Categorisation of these units classified the contributions at four knowledge levels: (1) *observations* of patient signs, (2) *findings* referring to clinically significant clusters of observations, (3) *facets* referring to pathophysiological states or broad categories of disease, and (4) *diagnoses* referring to clinical conclusions. They also categorised three *types of decisions* at the level of: findings, actions to be taken in patient management, and assessments of the overall state. The results showed how contributions were distributed over the different participants, how content changed over three consecutive days of patient care, how interactions, content, and reasoning differed in two types of intensive care units, and what episodes were particularly useful for expertise development. The data were both qualitatively and quantitatively described. This research in the tradition of naturalistic decision making (Klein, 2008) provides a good example of how expertise can be examined when it is distributed over multiple agents in a complex, real-life, dynamic environment.

Two other studies using this method have analysed the discussions in learning situations to examine the application of biomedical and clinical knowledge in problem-based learning tutorials with real-patients (Diemers, van de Wiel, Scherpbier, Heineman, & Dolmans, 2011), and the topics discussed in tutorial dialogues on diagnostic reasoning of trainees and their supervisors in general practice (Stolper et al., 2015). In the study conducted by Diemers and colleagues, a purposive sample of tutorial group discussions was divided into a preparation and a reporting phase. Based on a technique of proposition analysis for medical protocols (Patel & Groen, 1986), all transcripts were segmented into small meaningful information units or propositions. Propositions connect two concepts by a qualifier, such as “a pseudo polyp *is characteristic of* ulcerative colitis”. Guided by previous research and pilot interviews, we coded the propositions as *patient information, formal clinical knowledge, biomedical knowledge, informal clinical knowledge, procedural information,* and *other information*, and also indicated if the proposition was put forward by the tutor or a student. In this way, we could compare both the number of propositions per coding category, and the number of propositions contributed by tutors and students in both phases. To analyse the function of biomedical knowledge, we categorised what the biomedical lines of reasoning in the protocols explained, and found that

they mostly linked underlying mechanisms of disease to clinical features of patients, as intended by the educational format. In the study conducted by Stolper and colleagues, a representative sample of tutorial dialogues was taken and segmented based on turns in the conversation and also on content changes. The coding proceeded in a bottom-up and iterative way, but was informed by the researchers knowledge of diagnostic reasoning and research goals, which helped to characterise the topics of discussion. As trainees usually presented several patient cases for discussion, we differentiated between a reporting and an analysis phase. Segments were double-coded to indicate the contributions of trainees and supervisors. The number of words per code was counted so that we could examine to what extent the different topics were discussed and by whom. In line with the research questions, the data for all main coding categories were reported in tables, and specified in more detail for the categories of diagnostic reasoning and gut feelings. In addition, the tutorial dialogues, diagnostic reasoning, and the way in which gut feelings featured in the dialogues were described in qualitative terms. In both studies, the methods of analysis yielded rich but precise data that could be used to quantitatively compare the attention paid by the participant groups to the coding categories and the variables of interest, and qualitatively describe, interpret and illustrate these variables.

5.3 Free recall protocols

Free recall is a classical method used to study problem representation underlying task performance (Chi, 2006b; Chi et al., 1988; Feltovich et al., 2006). The most well-known example of a free recall study is the work on chess expertise of de Groot (1946/1978), who asked participants to think of a move in an actual game position before they had to recall the position. Chess masters selected the best moves and recalled more chess pieces than other players. This was explained as being the result of their better overall conception of the problem. Masters could grasp the problem at a high level in a very short time showing their expert knowledge of chess. The memory paradigm used, i.e., asking the chess players to recall briefly presented chess positions, has since been adopted to chart the problem representations and underlying knowledge structures by reviewing how individual chess pieces are chunked together in memory (Chase & Simon, 1973). The rationale behind the free recall measure is that the content of working memory is retrieved immediately after task performance (see Figure 1). Therefore, matching the goals of processing in the experimental situation to the real task is a prerequisite consideration in the development of task instructions (Ericsson & Smith, 1991; Ericsson, Patel & Kintsch, 2000). This is underlined by levels of processing theory showing that meaningful processing gives the best results due to the richer connections in memory (Craik, 2002).

The use of the free recall method in medical expertise research clearly shows that results are influenced by task processing conditions. In contrast to the superiority of expert memory for meaningful materials found in a wide variety of domains (Chi, 2006b; Feltovich et al., 2006), in medicine, findings have been less straightforward (Norman, et al., 2006). In most circumstances, experienced physicians appear to represent patient cases in a more condensed way than advanced students, if they process them for diagnosis (Schmidt & Boshuizen, 1993; de Bruin, van de Wiel, Rikers, & Schmidt, 2005). A sequence of recall studies manipulating case materials and task instructions suggests that clinical case processing of experts using patient descriptions is rather robust. However, memorisation instructions, perceiving the task as a memory task, or instructions for elaborate processing may enhance their recall (de Bruin, et al., 2005; van de Wiel, Schmidt, & Boshuizen, 1998; van de Wiel, Ploegh, Boshuizen, & Schmidt, 2005; Wimmers, Schmidt, Verkoeijen, & van de Wiel, 2005). If lab data were processed without further patient information and under elaborate problem formulation conditions, experts outperformed students in recalling the lab data after this analytical diagnostic task (Norman et al., 1989; Wimmers et al., 2005). If medical students knew in advance that they would be asked to recall a patient case (intentional recall condition) they recalled more case information than if they did not know this (incidental recall condition) (van de Wiel et al., 2005). In conclusion, the method has some drawbacks that are hard to control in experimental research, if only because more than one patient case should be presented. A lesson learned is that diagnostic tasks should be presented in a realistic way by aligning processing goals and providing the patient information that would be available

in practice. For research in visual expertise this means that both the image and the information physicians have before interpreting the image should be presented (Hatala, Norman, & Brooks, 1999; Kulatunga-Moruzi, Brooks, & Norman, 2004). To capture the nature of expertise, the information provided in patient case descriptions needs to be presented in a standard order and phrased as it is communicated in practice, either in the words of patients or of colleagues, and interpretation must be left to the participants. The manipulation of task performance conditions can be an important strategy to further uncover expert knowledge structures. A good example is constraining the time in which participants have to process case materials, as it might be expected that this will have a lower impact on experts than on less advanced participants (Schmidt & Boshuizen, 1993; van de Wiel et al., 1998). In addition, the method of free recall should be supplemented with other verbal protocols in order to corroborate findings. Representations of patient cases, for example, might be investigated by asking physicians how they would summarise or characterise a case or describe what information was critical for their diagnosis. In visual domains, recall protocols of the images presented can be obtained by asking participants to describe what they saw, and this can be supplemented with instructions to indicate the relevant features on the image, or draw the image (e.g., Lesgold et al., 1988; Gilhooly et al., 1997). In this way, the representation of features recognised can be separated from the interpretation of the pattern of features in diagnosis.

Analysis of free recall protocols, characteristic summaries, or protocols with critical cues can proceed in a straightforward way using proposition analysis. The number of propositions in the protocol that match the propositions in the case materials is counted. The number of summaries, i.e., inferences referring to more than one case proposition, may be counted separately to show to what extent participants represent the case information at a higher interpretative level. An example of a summary encompassing four propositions is “Auscultation reveals mitral valve insufficiency”, which summarises the more detailed information, “Auscultation reveals a holosystolic murmur at the apex radiating towards the axilla” (van de Wiel et al., 1998). The summaries can be provided at different levels of detail, varying from interpretation of lab data to the encapsulation of data into pathophysiological mechanisms or diagnostic labels. Moreover, the order of the information recalled may also reveal how the problem and underlying knowledge are represented in memory (e.g., Boshuizen & Claessen, 1985). The research questions will determine what to focus on in designing the study and analysing the data.

5.4 Explanation protocols

In medical expertise research, the post-hoc explanation method was introduced by Feltovich and Barrows (1984), and further developed by Patel and Groen (1986) to gain insight into the knowledge used in the diagnostic process and the causal lines of reasoning. Participants are asked to provide a pathophysiological explanation of the signs and symptoms in a patient case that they have processed for diagnosis. Just as in free recall, it is assumed that the knowledge activated in case processing will be retrieved when providing the explanation (see Figure 1). Although, in general, this seems to be the case for expert processing, explanations may be influenced by the knowledge available to understand the case materials. As the knowledge of novices and intermediates, for example, is fragmented and less coherently organised, they tend to elaborate on their knowledge when explaining the recalled case features (Boshuizen & Schmidt, 1992). It is clear, however, that explanations of patient cases can reveal the knowledge participants use in linking case information, pathophysiological mechanisms, and disease. Explanations allow for analysis of both the content and the structure of knowledge, i.e., how knowledge elements are related (Chi, 1997).

The explanation method is easy to use, but may be labour-intensive to analyse. Based on two studies in which we applied this method to compare the knowledge structures of medical students with those of experienced physicians (van de Wiel et al., 2000), and to examine the knowledge development of medical students over the period of their course (Diemers, et al., 2015), a detailed account of how to proceed with analysis and reporting are provided. For these research goals of comparing knowledge between groups and over time, the first step in analysis is to develop a model explanation. A model explanation links the signs

and symptoms in a case to the diagnosis via the most important biomedical and clinical concepts explicating the disease processes in a network representation much like a concept map. The model explanation represents a causal model of a disease and must be developed in close collaboration with experts. Concept mapping is a particularly useful tool to elicit experts' knowledge for this purpose (Hoffman & Lintern, 2006; Shadbolt & Smart, 2015). The participants' written explanation protocols are then translated into networks of linked concepts and compared to the model explanation network. The network representations enable an assessment of *explanation quality* in terms of correctness of both concepts and links. In our studies, the variables we coded in the protocols included the total *number of concepts* used in explanations, specified as the number of *model concepts, alternative concepts, detailed concepts,* and *wrong concepts*, as well as the total *number of links*, specified as *model links, alternative links, detailed links, wrong links* and *shortcuts* in reasoning. As measures of quality, we used the *percentage of model concepts* (relative to the total number of concepts used in the protocols) and the *percentage of model links* (relative to the total number of links used in the protocols). In the Diemers et al. study, we also counted the number of biomedical and clinical concepts used. The outcomes of the variables were depicted in graphics or tables to support interpretation. The method provides detailed insight into the quality of knowledge structures, (causal) lines of reasoning, and the nature of the knowledge used. Moreover, it allows meaningful comparisons between groups and conditions. The combination of variables allows consistent patterns to be found in the data. Explanation protocols provide a rich source of clear examples of expert and flawed knowledge and reasoning.

Two studies asking for explanations in the visual domain demonstrate the use of a keyword technique to characterise the type of utterances made by participants (Gilhooly et al., 1997; Jaarsma, Jarodzka, Nap, van Merriënboer, & Boshuizen , 2014). Gilhooly and colleagues were interested in whether cardiologists had more biomedical knowledge available to explain ECG traces than less experienced participants. Participants could look at the the ECG traces while providing the explanations. A computer program analysed how many words in the protocols indicated *trace characteristics, clinical inferences,* and *biomedical inferences* (similar to the analysis of the ECG think-aloud protocols that were collected a week before the explanations). This procedure resulted in very clear data which revealed the expected expertise effect. In the study conducted by Jaarsma and colleagues, participants were asked to give an explanation for their diagnosis after they had seen a microscopic image for two seconds. Two coders categorised the words in the combined protocols of 20 explanations. Some of these categories were based on previous research and others emerged from the data. The number of words in each coding category characterised patterns of reasoning processes and the types of knowledge used by each of the three expertise groups. These examples show that explanation protocols can be reliably used to examine expertise differences in the visual domain, and that indicative words can be used as practical units of analysis.

Providing explanations can be a task in its own right (Chi, 1997), as participants can for example be asked to explain a concept. This use of explanation protocols overlaps with knowledge elicitation techniques. Explanations collected in medical expertise research can be analysed from a psychological perspective to reveal differences between expertise groups in terms of the content and organisation of their knowledge. For example, in a study on the explanation of clinical concepts, we analysed the elaborateness, quality, and fluency with which explanations were provided (van de Wiel, Boshuizen, Schmidt, & Schaper, 1999). The researchers told participants that they were interested in what they knew about certain concepts, and asked them to explain 20 concepts to the experimenter in approximately 2 minutes. The instructions were carefully crafted to ensure that participants communicated all they knew during the full time slot, indicated when they were not sure, and also explained how they could recognise a particular concept in patients. Throughout the experiment, the experimenter guided the elicitation process according to this procedure in order to increase data quality. For each concept, a model explanation was constructed. This highlighted model concepts and referred to a definition of the concept, the major causes and clinical consequences, and the essential pathophysiological mechanisms of disease. The model explanations were based on medical literature and checked by medical specialists. The participants' transcribed explanation protocols were segmented into meaningful information units that were then coded based on content. *Elaborateness of the protocols* was measured by counting the total number of medical concepts used and specified per category: definition,

cause, clinical knowledge, pathophysiological knowledge, and therapeutic knowledge. *Quality of the explanations* was measured by comparing the explanation protocols to the model explanation revealing the number of model concepts and imprecise expressions used, and the number of clinical concepts that were unknown. *Accessibility of knowledge* was operationalised by the fluency with which participants provided their explanations and measured by the number of times they abruptly changed the subject, used thinking pauses, stumbled, thought aloud, or referred to their lack of knowledge. Coding was very precise and laborious and, as a result, reliable. The results could be clearly depicted in a table and gave good insights into the availability and accessibility of knowledge in the three expertise groups. The data were also interpreted in a qualitative manner for medical education purposes to illustrate major misconceptions in concepts that were weakly explained. This method may be particularly useful to chart knowledge and misunderstandings in complex domains, including visual expertise (e.g., the feature description test as used by Kok et al., 2013), and to effectively design instruction. If explanations are requested in a written format, the method is more feasible and shows availability of knowledge but not the fluency by which it is accessed. The method can also be used for assessment of and feedback to students.

5.5 Retrospective reports

The questions asked to participants in retrospective reports of problem solving are similar to those that can be asked in interviews. Examples of such questions are: “How did you solve the problem?” or “What did you think while problem solving?”. The important difference is that the delay between problem solving and answering these questions is minimised in retrospective reports that are generated immediately after task performance. This delay can undermine the validity of the answers because people tend to present their thought processes as being more coherent and intelligent than they are, and reconstruct their memories of the problem-solving processes based on the outcomes (Ericsson & Simon, 1980, 1993; van Someren et al., 1994). The effect of delay may also exist in retrospective reports if the task lasts longer than 10 seconds, as, after that time, the sequence of thoughts is no longer readily available in working memory (Ericsson & Simon, 1993). Another threat to the validity of the data is if, during questioning, researchers probe participants to give post-hoc rationalisations of what they did, e.g., by asking: “Can you explain how you proceeded in solving the problem?”, “What general approach did you use in problem solving?” or “Why did you solve the problem in this way?” If, for example, participants did not use a general, structured approach, but rather solved the problem by trial and error, they may feel embarrassed to say so. However, when interviews are well-prepared and conducted, the interviewer can tap into relevant knowledge and strategies the participants may have used by emphasising that there are no right or wrong answers, encouraging participants to think aloud, and probing. As demonstrated by knowledge elicitation practices, researchers can scaffold participants in a collaborative process to articulate what they know, even if this has not been articulated before (Hoffman et al., 1995; Hoffman & Lintern, 2006). A good example of such a method is cued or stimulated recall, in which participants are walked through the task by watching and/or listening to a recording of the problem-solving process while expressing what they remember of their thoughts at specific points. This method may be particularly useful if concurrent thinking aloud is not possible because the incoming information that needs to be attended to is presented faster than it can be verbalised. The recording then provides cues in working memory that trigger retrieval of the cognitive processes and mental representations involved in task performance (see Figure 1). For this reason, the method might be well suited to examine processing and interpretation of visual data. To obtain valid data, retrospective questions or reports should be gathered in representative samples of participants and validated by the results of other methods, such as thinking aloud. To analyse the data, researchers can develop a coding scheme (just as in interviews or think-aloud protocols) and report the data in both qualitative and quantitative ways.

The following two examples will show how the method of stimulated recall can be effectively applied to gather additional data on participant’s thoughts while accomplishing a task. In a study on problem analysis in a tutorial group, the recorded discussion was presented to each individual group member immediately after the meeting to elicit their memories of their thinking during the discussion (de Grave,

Boshuizen, & Schmidt, 1996). Participants could stop the videotape at any time to recall what they had in mind while the others were talking. The research goal was to investigate whether problem-based learning leads to conceptual change when students participate in small group discussions. The data showed that this stimulated recall procedure provided further insight into the prior knowledge invoked, the changes in reasoning based on the group's theory building process, and the metacognitive reflections engaged in, as well as participants' thoughts about the group process. A coding scheme guided analysis of both the group interaction and the stimulated recall protocols, enabling comparisons between the number of clauses in each coding category across the two types of protocols. A temporal analysis of the categories of theory building and meta-reasoning showed how these thinking processes interacted and how conceptual change came about. This was illustrated with the stimulated recall protocol of one student. This example shows that stimulated recall can be a very informative means of elucidating covert thinking processes in group discussions. In the other example, retrospective reports of electrical circuit problem solving were elicited by presenting the participant's eye fixations and mouse-keyboard operations on the computer screen in the original task (van Gog, Paas, van Merriënboer, Witte, 2005). This may be a very good way of capturing the cognitive processing in perceptual-motor tasks. The goal of the study was to compare the results collected when the different methods of thinking aloud, retrospective reporting, and cued retrospective reporting were used. Participants were asked to tell what they were thinking during the task while watching the record of their eye movements and actions. Unfortunately, the record was replayed at the same speed as in task performance, and participants could not stop the recording to share their thoughts. This was probably the reason why cued retrospective reporting did not deliver more information than thinking aloud for three of four coding categories. Task performance can slow down during thinking aloud, but this was not possible during the cued retrospective reporting procedure used in this study. In conclusion, the examples show that stimulated recall may provide valuable information about the knowledge and reasoning involved in task performance when it is well-designed and implemented.

6. Conclusion

The methods of interviewing and collecting verbal protocols provide rich data to examine expertise and expertise development from different perspectives in all kinds of domains. Interviews are needed in the exploratory phase of research to gather information on the tasks and the problem situations to be investigated. They can also be used to obtain objective data to compare different expertise groups on targeted topics. In verbal protocol studies, knowledge, mental representations, and reasoning are examined in direct relation to representative tasks in which experts should outperform less advanced or experienced participants. The protocols collected as a result of thinking aloud, dialogues, and group discussions tap into the concurrent cognitive processes of experts and students during task performance, whereas verbal protocols of free recall, explanations, and retrospective reports are gathered after the task has been performed. All methods have their advantages and disadvantages in terms of the validity of the data obtained. To grasp the full nature of domain expertise, different methods should be applied in order to complement one another. Sometimes, different verbal protocols can be gathered in the same study, for example, thinking aloud while interpreting patient information and making a diagnosis, incidental recall of the last case presented, and explanation of the case materials a week later (e.g., Gilhooly et al., 1997). The selection of tasks and case materials is crucial and should be well-aligned with practices in real-life settings in order to capture expertise. Varying task difficulty is an important manipulation that can be used to investigate in what situations, and for which participant groups, automatic processing falls short and deliberate thinking is involved. The interactions between knowledge, cognitive processing, and task characteristics are key to the understanding of expertise. Interviews and cognitive task analysis play an important role in identifying the relevant interaction patterns, before research using verbal protocols can be designed. The formulation of interview questions and task instructions need careful attention in order to safeguard data validity when examining the underlying cognitions of expertise. Analysis of interviews and verbal protocols is usually

labour intensive, and reducing analysis to manageable proportions is best guided by the specific research goals. Quantifying the qualitative variables identified in developing the coding schemes provides an overview that helps the researchers to find, interpret, and report main patterns in the data. All methods described in this article may contribute to further developing different domains of expertise. They can also be used to examine visual expertise, as professionals are usually able to communicate what they see, what conclusions they come to, and what they plan to do when collaborating with colleagues and teaching students.

Keypoints

- The research questions, available theory, and previous findings are the starting point for coding schemes used to analyse interviews and verbal protocols.
- Interviews can be used to obtain relevant information in an objective way to compare groups with different levels of expertise.
- Interviews provide indispensable insight in the tasks and materials to be selected for research gathering verbal protocols.
- As expertise emerges from expert knowledge in interaction with the task and the problems to be solved, the selection of participants and materials as well as the design of task instructions are critical in preparing verbal protocol studies.
- Different methods of research, such as interviews and verbal protocols, need to be combined to fully grasp expertise in complex domains, such as visual expertise.

Acknowledgments

I am very grateful for the valuable feedback I received from my colleagues, Els Boshuizen and Fleurie Nievelstein, and two anonymous reviewers on earlier drafts of this article.

References

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*(4), 355. doi:10.1037/0003-066X.51.4.355

Bartram, D. (2008). Work profiling and job analysis. In N. Chmiel (Ed.), *An introduction to work and organizational psychology: A European perspective* (2nd ed.) (pp. 3-28). Oxford, UK: Blackwell.

Brooks, J., McCluskey, S., Turley, E., & King, N. (2015). The utility of template analysis in qualitative psychology research. *Qualitative Research in Psychology, 12*(2), 202-222. doi:10.1080/14780887.2014.955224.

Boshuizen, H. P. A., & van de Wiel, M. W. J. (1998). Multiple representations in medicine: How students struggle with it. In M. W. van Someren, P. Reimann, H. P. A. Boshuizen, & T. de Jong (Eds.). *Learning with multiple representations.* Amsterdam, the Netherlands: Elsevier.

Boshuizen, H. P. A., van de Wiel, M. W. J., & Schmidt, H. G. (2012). What and how advanced medical students learn from reasoning through multiple cases. *Instructional Science,* 40(5), 755-768. doi:10.1007/s11251-012-9211-z.

Boshuizen, H. P. A., & van de Wiel, M. W. J. (2014). Expertise development through schooling and work. In A. Littlejohn, & A. Margaryan (Eds.), *Technology-enhanced professional learning: Processes, practices and tools* (pp. 71-84). New York, NY: Taylor & Francis/Routledge.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55-81.

Chi, M. T. H., (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, *6*(3), 271-315. doi:10.1207/s15327809jls0603_1

Chi, M. T. H. (2006a). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 21-30). New York, NY: Cambridge University Press.

Chi, M. T. H. (2006b). Laboratory methods for assessing experts' and novices' knowledge. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 167-184). New York, NY: Cambridge University Press.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145-182. doi:10.1016/0364-0213(89)90002-5

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices, *Cognitive Science*, *5*, 121-152. doi:10.1207/s15516709cog0502_2

Chi, M. T. H., Glaser R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum.

Chipman, S. F., Schraagen, J. M., & Shalin, V. L. (2000). Introduction to cognitive task analysis. In J. M. Schraagen, S. F. Chipman, & V. L. Shalin (Eds.), *Cognitive task analysis* (pp. 3-23). Mahwah, NJ: Lawrence Erlbaum.

Claessen, H. F. A., & Boshuizen, H. P. A. (1985). Recall of medical information by students and doctors. *Medical Education*, *19*(1), 61-67. doi:10.1111/j.1365-2923.1985.tb01140.x

Craik, F. I. (2002). Levels of processing: Past, present... and future? *Memory*, *10*(5-6), 305-318. doi:10.1080/09658210244000135

Custers, E. J. (2013). Medical education and cognitive continuum theory: An alternative perspective on medical problem solving and clinical reasoning. *Academic Medicine*, *88*(8), 1074-1080. doi:10.1097/ACM.0b013e31829a3b10

Deakin, H., & Wakefield, K. (2014). Skype interviewing: Reflections of two PhD researchers. *Qualitative Research*, *14*(5), 603-616. doi:10.1177/1468794113488126

de Bruin, A. B. H., van de Wiel, M. W. J.,. Rikers, R. M. J. P, & Schmidt, H. G. (2005). Examining the stability of experts' clinical case processing: An experimental manipulation. *Instructional Science*, *33*, 251-270. doi:10.1007/s11251-005-3598-8

De Grave, W. S., Boshuizen, H. P. A., & Schmidt, H. G. (1996). Problem based learning: Cognitive and metacognitive processes during problem analysis. *Instructional Science*, *24*(5), 321-341. doi:10.1007/BF00118111

de Groot, A. D. (1978). *Thought and choice in chess. (2nd ed)*. The Hague, The Netherlands: Mouton. (Original work published in 1946)

de Leng, B., Dolmans, D., van de Wiel, M. W. J., Muijtjens, A., & van der Vleuten, C. (2007). How video cases should be used as authentic stimuli in problem-based medical education. *Medical Education*, *41*, 181-188. doi:10.1111/j.1365-2929.2006.02671.x

Diemers, A. D., van de Wiel, M. W. J., Scherpbier, A. J., Baarveld, F., & Dolmans, D. H. (2015). Diagnostic reasoning and underlying knowledge of students with preclinical patient contacts in PBL. *Medical education*, *49*(12), 1229-1238. doi:10.1111/medu.12886.

Diemers, A. D., van de Wiel, M. W. J, Heineman, E., Scherpbier, A. J. J. A., & Dolmans D. H. J. M. (2011). Pre-clinical patient contacts and the application of biomedical and clinical knowledge. *Medical Education*, *45*, 280-288. doi:10.1111/j.1365-2923.2010.03861.x.

DuBois, D., & Shalin, V. L. (2000). Describing job expertise using cognitively oriented task analyses (COTA). In J. M. Schraagen, S. F. Chipman, & V. L. Shalin (Eds.), *Cognitive task analysis* (pp. 41-55). Mahwah, NJ: Lawrence Erlbaum.

Elstein, A. S., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *British Medical Journal*, *324*, 729-732.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Emans, B. (2004). *Interviewing: Theory, techniques and training*. Groningen, The Netherlands: Wolters-Noordhoff.

Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. Mahwah, NJ: Lawrence Erlbaum.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*(10 Suppl), S70-81. doi:00001888-200410001-00022

Ericsson, K. A. (2006a). An introduction to the Cambridge handbook of expertise and expert performance: Its development, organization, and content. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 3-19). New York, NY: Cambridge University Press.

Ericsson, K. A. (2006b). Protocol analysis and expert thought: Concurrent verbalisations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 223-241). New York, NY: Cambridge University Press.

Ericsson, K. A. (2009). *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments.* New York, NY: Cambridge University Press.

Ericsson, K. A. (2015). Acquisition and maintenance of medical expertise: A perspective from the expert-performance approach with deliberate practice. *Academic Medicine, 90*(11), 1471-1486. doi:10.1097/ACM.0000000000000939

Ericsson, K. A. (2014). How to gain the benefits of the expert performance approach in domains where the correctness of decisions are not readily available: A reply to Weiss and Shanteau. *Applied Cognitive Psychology, 28*(4), 458-463. doi:10.1002/acp.3029

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245. doi:10.1037/0033-295X.102.2.211

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406. doi:10.1037/0033-295X.100.3.363

Ericsson, K. A., Patel, V., & Kintsch, W. (2000). How experts' adaptations to representative task demands account for the expertise effect in memory recall: Comment on Vicente and Wang (1998). *Psychological Review, 107*(3), 578-592. doi:10.1037/0033-295X.107.3.578

Ericsson, K. A., & R. Poole (2016). *Peak: Secrets from the new science of expertise.* London, UK: The Bodley Head.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-251. doi:10.1037/0033-295X.87.3.215

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: An introduction. In K. A. Ericsson, & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 1-38). Cambridge: Cambridge University Press.

Evetts, J., Mieg, H. A., & Felt, U. (2006). Professionalization, scientific expertise, and elitsm: A sociological perspective. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 105-123). New York, NY: Cambridge University Press.

Feltovich, P. J., & Barrows, H. S. (1984). Issues of generality in medical problem solving. In H. G. Schmidt, & M. L. De Volder (Eds.), *Tutorials in problem-based learning. New directions in training for the health professions*, (pp. 128-142). Assen/Maastricht: Van Gorcum.

Feltovich, P. J., Prietula, M. J., & Ericsson, A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 41-67). New York, NY: Cambridge University Press.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*(4), 327. doi:10.1037/h0061470

Gegenfurtner, A., Siewiorek, A., Lehtinen, E., & Säljö, R. (2013). Assessing the quality of expertise differences in the comprehension of medical visualizations. *Vocations and Learning*, *6*(1), 37-54. doi:10.1007/s12186-012-9088-7.

Gilhooly, K. J., McGeorge, P., Hunter, J., Rawles, J. M., Kirby, I. K., Green, C., & Wynn, V. (1997). Biomedical knowledge in diagnostic thinking: The case of electrocardiogram (ECG) interpretation, *European Journal of Cognitive Psychology*, *9*(2), 199-223. doi:10.1080/713752555.

Hamm, R. M. (1988). Clinical intuition and clinical analysis: Expertise and the cognitive continuum. In J. Dowie, & A. Elstein (Eds.), *Professional judgment: A reader in clinical decision making*, (pp.78-105). Cambridge, MA: Cambridge University Press.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *Transactions on Systems, Man and Cybernetics, IEEE*, *17*(5), 753-770. doi:10.1109/TSMC.1987.6499282

Hashem, A., Chi, M. T., & Friedman, C. P. (2003). Medical errors as a result of specialization. *Journal of Biomedical Informatics*, *36*(1), 61-69. doi:10.1016/S1532-0464(03)00057-1

Hassebrock, F., & Prietula, M. J. (1992). A protocol-based coding scheme for the analysis of medical reasoning. *International Journal of Man-Machine Studies*, *37*(5), 613-652. doi:10.1016/0020-7373(92)90026-H.

Hatala, R., Norman, G. R., & Brooks, L. R. (1999). Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *Journal of General Internal Medicine*, *14*(2), 126-129. doi:10.1111/j.1525-1497.1999.tb00008.x

Helle, L. (2017). Prospects and pitfalls in combining eye-tracking data and verbal reports. *Frontline Learning Research*, *5*(3), 1-12. doi:10.14786/flr.v5i3.254

Hoffman, R. R., & Lintern, G. (2006). Eliciting and representing the knowledge of experts. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 203-222). New York, NY: Cambridge University Press.

Hoffman, R. R., Shadbolt, N. R., Burton, A. M., & Klein, G. (1995). Eliciting knowledge from experts: A methodological analysis. *Organizational behavior and human decision processes*, *62*(2), 129-158.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277-1288. doi:10.1177/1049732305276687

Jaarsma, T., Jarodzka, H., Nap, M., Merrienboer, J. J. G., & Boshuizen, H. P. A. (2014). Expertise under the microscope: Processing histopathological slides. *Medical Education*, *48*(3), 292-300. doi:10.1111/medu.12385.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515-526. doi:10.1037/a0016755

King, N., & Horrocks, C. (2010). *Interviews in qualitative research*. London, UK: Sage.

Klein, G. (2008). Naturalistic decision making. *Human Factors*, *50*(3), 456-460. doi:10.1518/001872008X288385.

Kok, E. M., Jarodzka, H., de Bruin, A. B. H., BinAmir, H. A. N., Robben, S. G. F., & van Merriënboer, J. J. G. (2015). Systematic viewing in radiology: Seeing more, missing less?. *Advances in Health Sciences Education*, 1-17. doi:10.1007/s10459-015-9624-y.

Kok, E. M., de Bruin, A. B. H., Robben, S. G. F., & van Merriënboer, J. J. G. (2013). Learning radiological appearances of diseases: Does comparison help? *Learning and Instruction*, *23*, 90-97. doi:10.1016/j.learninstruc.2012.07.004.

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology (3rd ed.)*. Newsbury Park, CA: Sage.

Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research (5th edition)*. Thousand Oaks, CA: Sage.

Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2004). Using comprehensive feature lists to bias medical diagnosis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 563-572. doi:10.1037/0278-7393.30.3.563

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., Wang, Y., et al. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 311-342). Hillsdale, NJ: Lawrence Erlbaum.

Lesgold, A. (2000). On the future of cognitive task analysis. In J. M. Schraagen, S. F. Chipman, & V. L. Shalin (Eds.), *Cognitive task analysis* (pp. 451-465). Mahwah, NJ: Lawrence Erlbaum.

Mieg, H. A. (2006). Social and sociological factors in the development of expertise. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 743-760). New York, NY: Cambridge University Press.

Morgan, D. L. (1996). *Focus groups as qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Neufeld, V. R., Norman, G. R., Feightner, J. W., & Barrows, H. S. (1981). Clinical problem-solving by medical students: A cross−sectional and longitudinal analysis. *Medical Education*, *15*(5), 315-322. doi:10.1111/j.1365-2923.1981.tb02495.x

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1166-1174. doi:10.1037/0278-7393.15.6.1166

Norman, G. R., Eva, K., Brooks, L., & Hamstra, S. (2006). Expertise in medicine and surgery. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 339-354). New York, NY: Cambridge University Press.

Patel, V. L., & Arocha, J. F. (2001). The nature of constraints on collaborative decision making in health care settings. In E. Salas, & Klein, G. (Eds.), *Linking expertise and naturalistic decision making* (pp. 383-405). Mahwah, NJ: Lawrence Erlbaum.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science*, *10*(1), 91-116. doi:10.1207/s15516709cog1001_4

Patel, V. L., Kaufman, D. R., & Magder, S. A. (1996). The acquisition of medical expertise in complex dynamic environments. In K .A. Ericcson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*, (pp. 127-165). Mahwah, NJ: Lawrence Erlbaum.

Prince, K. J. A. H., van de Wiel, M. W. J., van der Vleuten, C. P. M., Boshuizen, H. P. A., & Scherpbier, A. J. J. A. (2004). Junior doctors' opinions about the transition from medical school to clinical practice: A change of environment. *Education for Health*, *17*(3), 323-331. doi:10.1080/13576280400002510

Salas, E., & Klein, G. A. (Eds .) (2001). *Linking expertise and naturalistic decision making*. Mahwah, NJ: Lawrence Erlbaum.

Salas, E., Rosen, M. A., Burke, C. S., Goodwin, G. F., & Fiore, S. M. (2006). The making of a dream team: When expert teams do best. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 439-453). Cambridge, UK: Cambridge University Press.

Sanchez, J. I., & Levine, E. L. (2012). The rise and fall of job analysis and the future of work analysis. *Annual Review of Psychology*, *63*, 397-425. doi: 10.1146/annurev-psych-120710-100401.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On the origin of intermediate effects in clinical case recall. *Memory and Cognition*, *21*, 338 - 351. doi:10.3758/BF03208266

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, *53*(2), 252-266.

Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, *136*(2), 253-263.

Skopec, E. W. (1986). *Situational interviewing*. Prospects Heights, II: Waveland Press.

Stalmeijer, R. E., McNaughton, N., & Van Mook, W. N. K. A. (2014). Using focus groups in medical education research: AMEE Guide No. 91. *Medical Teacher*, *36*(11), 923-939. doi:10.3109/0142159X.2014.917165

Stolper, E., van Bokhoven, M., Houben, P., Van Royen, P., van de Wiel, M. W. J., van der Weijden, T., & Dinant, G. J. (2009). The diagnostic role of gut feelings in general practice. A focus group study of the concept and its determinants. *BMC Family Practice, 10*(1), 17. doi:10.1186/1471-2296-10-17.

Stolper, E., van de Wiel, M. W. J., Hendriks, R. H. M., Van Royen, P., Van Bokhoven, M., Van der Weijden, T., & Dinant, G. J. (2015). How do gut feelings feature in tutorial dialogues on diagnostic reasoning in GP traineeship? *Advances in Health Sciences Education, 20,* 499-513. doi:10.1007/s10459-014-9543-3.

Stolper, E., van de Wiel, M. W. J., van Bokhoven, M., Van Royen, P., van der Weijden, T., & Dinant, G. J. (2011). Gut feelings as a third track in general practitioners' diagnostic reasoning. *Journal of General Internal Medicine, 26,* 197-203. doi:10.1007/s11606-010-1524-5.

Shadbolt, N. R., & Smart, P. R. (2015) Knowledge elicitation. In J. R. Wilson, & S. Sharples (Eds.), *Evaluation of Human Work* (4th ed.). Boca Raton, FL: CRC Press.

Tracey, T. J., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist*, *69*(3), 218- 229. doi:10.1037/a0035099

van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (2000). Knowledge restructuring in expertise development: Evidence from pathophysiological representations of clinical cases by students and physicians. *European Journal of Cognitive Psychology, 12*(3), 323-355. doi:10.1080/09541440050114543

van de Wiel, M. W. J., Boshuizen, H. P. A., Schmidt, H. G. & Schaper, N. C. (1999). The explanation of medical concepts by expert physicians, clerks and advanced students. *Teaching and Learning in Medicine, 11*(3), 153-163. doi:10.1207/S15328015TL110306

van de Wiel, M. W. J., Ploegh, K., Boshuizen, H. P. A., & Schmidt, H. G. (2005). *The influence of diagnosis and memorization instructions on clinical case processing by students and physicians.* Paper presented at the Annual Meeting of the American Educational Research Association 2005. Montreal, Canada, April 11-15.

van de Wiel, M. W. J., Schaper, N. C., Scherpbier, A. J. J. A., Van der Vleuten, C. P. M., & Boshuizen, H. P. A. (1999). Students' experiences with real patient tutorials in a problem-based curriculum. *Teaching and Learning in Medicine, 11*(1), 12-20. doi:10.1207/S15328015TLM1101_5

van de Wiel, M. W. J., & Schmidt, H. G., Boshuizen, H. P. A. (1998). A failure to reproduce the intermediate effect in clinical case recall. *Academic Medicine, 73*(8), 894-900.

van de Wiel, M. W. J., & Van den Bossche, P. (2013). Deliberate practice in medicine: The motivation to engage in work-related learning and its contribution to expertise. *Vocations and Learning, 6*(1), 135-158. doi:10.1007/s12186-012-9085-x.

van de Wiel, M. W. J., Van den Bossche, P., Janssen, S., & Jossberger, H. (2011). Exploring deliberate practice in medicine: How do physicians learn in the workplace? *Advances in Health Sciences Education. 16*(1), 81-95. doi:10.1007/s10459-010-9246-3.

van de Wiel, M. W. J., Van den Bossche, P., & Koopmans, R. P. (2011). Deliberate practice, the high road to expertise: K.A. Ericsson. In Dochy, F., Gijbels, D., Segers, M., & Van den Bossche, P. (Eds.), *Theories of learning for the workplace: Building blocks for training and professional development programs* (pp. 1-16). London, UK: Routledge.

van Gog, T., Paas, F., Van Merriënboer, J. J., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied, 11*(4), 237-244. doi:10.1037/1076-898X.11.4.237

van Someren, M. V., Barnard, Y. F., & Sandberg, J. A. (1994). *The think aloud method: A practical approach to modelling cognitive processes*. London, UK: Academic Press.

Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. *Human Factors*, *45*(1), 104-116.

Wimmers, P. F., Schmidt, H. G., Verkoeijen, P. P. J. L., & van de Wiel, M. W. J. (2005). Inducing expertise effects in clinical case recall through the manipulation of processing. *Medical Education, 39*, 949-957. doi:10.1111/j.1365-2929.2005.02250.x