Use of head camera-cued recall and debrief to externalise expertise: a systematic review of literature from multiple fields of practice.

Journal:	*BMJ Simulation & Technology Enhanced Learning*
Manuscript ID	bmjstel-2018-000341
Article Type:	Systematic review
Date Submitted by the Author:	17-Apr-2018
Complete List of Authors:	Blackhall, Vivienne; Highland Academic Surgical Unit, Raigmore Hospital and Centre for Health Science; Aberdeen University, Centre for Health Education Research and Innovation Walker, Kenneth; Raigmore Hospital, Dept of General Surgery Whiteley, Iya; University College London Department of Space and Climate Physics, Centre for Space Medicine Wilson, Philiip; University of Aberdeen Institute of Applied Health Sciences, Centre for Rural Health
Keywords:	recall debrief, decision making, expert, skills

Use of head camera-cued recall and debrief to externalise expertise: a systematic review of literature from multiple fields of practice.

Registration:
Our systematic review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) on 6th of March 2017 (registration number CRD42017057484).

Authors:
Corresponding author: Vivienne I Blackhall [a,b]
vblackhall@nhs.net
Other authors:
Kenneth G Walker [a,b]
Iya Whiteley [c]
Philip Wilson[d]

Author Affiliations:
a) Centre for Health Education Research and Innovation, College of Life Sciences and Medicine, Aberdeen University, Aberdeen AB24 3FX
b) NHS Highland Department of Medical Education, Centre for Health Science, Inverness IV2 3JH
c) Centre for Space Medicine, Mullard Space Science Laboratory, University College London, Holmbury St Mary, Surrey RH5 6NT
d) University of Aberdeen Centre for Rural Health, Centre for Health Science, Inverness IV2 3JH

Contributions:
PW is the guarantor. VB drafted the manuscript. All authors contributed to the genesis of the selection criteria. VB and PW developed the search strategy. VB, PW and KW screened the papers for inclusion. All authors read, gave feedback and approved the final version of the manuscript.

Competing interest:
IW authored two of the papers under review, under the names I Solodilova and I Solodilova-Whiteley.

Funding:
No funding was received for this work.

Support:
Database expertise was provided by Rob Polson, senior librarian at the Centre for Health Science Inverness.

Word count:
3409

ABSTRACT

Background: The study of decision making in complex naturalistic environments poses several challenges. In response to these, video-stimulated cued-recall-debrief was developed. It involves an individual wearing a head-mounted camera which records a task from their point-of-view. Afterwards, footage captured is reviewed along with a facilitated debrief to help externalise cognitive processes. In theory, motion, audio and visual cues generate a high level of experiential immersion which helps the expert to articulate previously hidden thoughts and actions.

Objective: To examine the current evidence for video-stimulated cued-recall-debrief as a means of explicating expert thoughts and feelings in complex tasks in a range of environments.

Study Selection: MEDLINE, EMBASE, ERIC, Sportdiscus, PsychINFO and Google Scholar were searched for articles containing the key terms 'cued-recall (debrief)', 'decision making', 'skills' and 'video recording'. Studies were included if they examined the following outcomes: i) feasibility, ii) extent of experiential immersion, iii) ability to generate unique insight into decision making processes, iv) current applications. 1831 articles were identified initially, nine studies were included in the final review.

Findings: Video-stimulated cued-recall-debrief is associated with a high level of experiential immersion and generates between two and four times the number of recollections compared with free recall. It can be used to build models of cognitive activity and to characterise the way in which more and less skilled individuals tend to think and feel.

Conclusions: The technique could be used to explicate expertise within medicine: these insights into performance could be used as a learning tool for trainees.

Keywords: Recall-debrief, decision making, expert, skills

INTRODUCTION

The study of naturalistic decision making in complex environments poses several problems for the researcher[1]. Whilst concurrent reporting can interrupt the flow of a procedure and alter the cognitive processes being studied, retrospective reporting, if not cued by real events, can be associated with bias and omission of key information. In response to these limitations, a technique known as video stimulated cued-recall-debrief was developed by Omodei et al[1].

The technique is designed to explicate the thoughts and actions of an individual undertaking a complex task. The individual wears a head-mounted camera which records the task from their own point-of-view. Footage captured is reviewed after the event along with a facilitated debrief to help externalise cognitive processes. The theory posits that footage from an 'own point-of-view perspective' contains motion, auditory and visual cues, which together stimulate memory processes, allowing the individual to undergo a high level of experiential immersion. This facilitates recall of specific cognitive processes as well as "pre-verbal" (or non-verbal) processes such as affect and motivation. Experiential immersion is important because recollection retrieval is superior in situations which more closely resemble the original event's context[1].

The method has been described in several domains including firefighting, orienteering, clinical medicine and the aviation industry. The aim of this systematic review is to examine the current evidence for video stimulated cued-recall-debrief as a means of explicating expert thoughts and feelings in complex tasks. The scope was intentionally not limited to a medical environment as the technique was initially described in other domains such as orienteering and firefighting. Understanding its application in these contexts will allow consideration of the potential range of uses of the technique.

The primary outcomes of interest are whether video stimulated cued-recall-debrief is:

- feasible (in terms of general practicability and extent of interference by the camera)
- associated with experiential immersion (and if so, to what degree)
- able to generate unique insight to the performer's decision making processes.

In addition, an evaluation of the current applications of the technique is offered.

METHODS

Data collection

A systematic literature search was conducted with the assistance of one librarian at our institution. Key terms were searched for in MEDLINE, EMBASE, ERIC, Sportdiscus, PsychINFO and Google Scholar from first records to April 2017. The literature search was limited to the English language and human subjects. Following database interrogation and identification of the relevant studies, the reviewers checked for additional relevant cited and citing articles, by scanning paper titles. The search terms were: 'cued-recall (debrief)', 'decision making', 'skills', 'video recording' or variations thereof. A detailed search strategy is provided in table 1.

1	exp decision making/ or exp clinical decision making/ or exp decision making, organizational/ or exp Decision making, computer assisted/
2	decision*.af.
3	cognitive process*.af.
4	thought process*.af.
5	choice*.af.
6	judgement*.af.
7	1 or 2 or 3 or 4 or 5 or 6
8	task*.af.
9	skill*.af.
10	psychomotor perform*.af.
11	procedure*.af.
12	operation*.af.
13	8 or 9 or 10 or 11 or 12
14	video*.af.
15	exp video-audio media/ or video recording/
16	head mounted camera*.af.
17	head camera*.af.
18	14 or 15 or 16 or 17
19	cue*.af.
20	Mental Recall/ or Memory/
21	memor*.af.
22	debrief*.af.
23	feedback*.af.
24	19 or 20 or 21 or 22 or 23
25	7 and 13 and 18 and 24

Table 1: MEDLINE search strategy. Of note cued recall* debrief.af yielded zero results and was therefore excluded. Advice sought from the librarian suggested that terms such as surgery (truncated to surg*) and education (truncated to educat*) should also be searched for in terms of floating subheadings (.fs). This yielded no additional papers to those retrieved by searching for 'all fields (.af)'

Literature search results were uploaded to the reference manager Mendeley. After removing duplicate studies, the team developed and tested screening questions based on the inclusion criteria. VB independently screened the titles and abstracts of manuscripts against the inclusion criteria. When there was dubiety as to the relevance of the article, the full text was also reviewed. The full text for all titles that appeared to meet the inclusion criteria were obtained (n=20). A flow chart displaying the selection process is provided in Figure 1. The full texts were formally assessed for eligibility in relation to the inclusion criteria. This was undertaken by three researchers. VB reviewed all texts. The manuscripts were arbitrarily divided into two and reviewed by PW and KW respectively. One hundred percent agreement was established between the reviewers. The researchers were not blinded to the journal titles, study authors nor the institutions.

All types of study design were included. This decision was made following an initial scoping search, where it became clear that the literature base was small. To our knowledge (and following interrogation of trials registries), there are no randomized controlled trials examining the utility of the 'cued-recall-debrief' technique.

Papers were included if the participants were considered to be experts, undertaking any complex task. No limit was placed upon the type of task and could, for example, include playing a tennis match. The search included all studies describing video stimulated cued-recall-debrief, or indeed any other similar feedback process of another name which is designed to elicit detailed recall in order to explicate expert's thoughts and feelings after the task. Comparators were all or none. Studies were included if they examined one or more of the following features of the technique:

- Feasibility
- Extent of experiential immersion
- Ability to generate unique insight into the performer's decision making processes
- Current applications

Quality assessment

All studies had a qualitative component. To assess the quality of the papers, Critical Appraisal Skills Programme (CASP) checklists were used[2]. There are no other published reporting guidelines which include a tool for assessing qualitative research in medical education. Manuscripts were not graded in terms of their quality, as the CASP checklist guidance notes state the checklists are designed to be used as educational tools, not scoring systems. The results were tabulated and areas of concern highlighted using a colour coding system (table 2). For mixed methods studies, quality assessment of quantitative components was included in a separate column.

	Unsworth	Unsworth	Unsworth	Pellacia	Omodei
Clear statement of aims	Yes. Description of techniques to study naturalistic decision making (NDM) and case report cued-recall-debrief	Yes. To conceptualise clinical reasoning of OTs.	Yes. To determine differences in clinical reasoning between novice and expert OTs	Yes, to identify how emergency medicine doctors make diagnoses	Yes, to conceptualise decision making of firefighters
Qualitative methodology appropriate?	Yes	Yes	Yes. Mixed methods.	Yes.	Yes. Mixed methods.
Research design appropriate?	Yes. Basic description	Yes	Yes.	Yes.	Yes, but basic description.
Recruitment strategy appropriate?	Can't tell– selection and recruitment process unclear.	Yes: participants nominated by OT manager and clients nominated by OTs.	Yes. Clear inclusion criteria. Clinicians nominated a client for inclusion.	Yes. Multicentre recruitment. 6 point inclusion criteria checklist.	Can't tell. No information provided.
Data collection appropriate?	Can't tell. No clear description of process	Yes, detailed description process.	Yes. Clear description. Identical prompts during debrief increases reliability	Yes. Detailed description of process.	Yes. Clear description.
Relationship between researcher & participants considered?	No critical examination of researcher's own role	Yes. Lead researcher did not conduct debriefs.	Yes. Researchers blinded to participant experience level.	Yes, researcher did not work in same hospital as participants.	No critical examination of researcher's own role.
Ethical issues considered?	No description of ethical approval, consent or confidentiality	Yes. Informed consent obtained. Ethical approval from University and clinical centre.	Yes. Ethical approval granted. Consent obtained.	Yes. Consent obtained from patients & experts. Ethics committee approval.	Can't tell. No details relating to ethics or consent.
Rigorous data analysis?	Anecdotal evidence. No formal analysis	Yes. Transcription of CRD commentary, *a priori* coding using established clinical reasoning models. 100% agreement between researchers in data coding. Thematic analysis.	Yes. Transcription and *a priori* coding CRD commentary. Coding discrepancies reviewed, subsequent near perfect agreement. Thematic analysis.	Yes. Transcription CRD commentary, deductive coding approach with development of codebook. Intercoder reliability 96%. Thematic analysis	Yes, *a priori* transcription and coding of CRD commentary with thematic analysis.
Quantitative methodology comments	NA	NA	Counts of types of codes. Chi^2 analyses to determine differences between groups.	NA	Counts of types of codes. Basic descriptive data.
Clear statement of findings?	Not particularly. Findings jumbled with existing literature. Difficult to identify novel data.	Yes. Identified types of reasoning employed and a new subset of clinical reasoning.	Yes. Clear differences in way in which experts and novices employ clinical reasoning.	Yes. Detailed description of when, how and number of diagnoses made.	Yes, novel finding re: importance emotional self regulation
Value of research	Novel study. Outlined 2 future planned studies	Yes, discusses contribution to what is already known and identifies new areas for research (broadening to other clinical settings).	Yes. Implications for promoting development of clinical reasoning skills amongst clinicians.	Yes. Discusses implications for training in order to improve diagnostic efficiency. Suggests avenues for future research.	Yes, suggests future direction for research (sets out aims of study 2). Implications for training firefighters.

	Omodei	Omodei	Elliot	Solodilova
Clear statement of aims	Yes. Determine cognitive processes associated with better versus poorer decision making in firefighters	Yes. To compare two methods for studying NDM during orienteering	Yes. To compare two methods for studying NDM in air defence simulations	Yes. To uncover how pilots gather and use information from their environment.
Qualitative methodology appropriate?	Yes. Mixed methods.	Yes. Mixed methods.	Yes. Mixed methods	Yes.
Research design appropriate?	Yes, but minimal description of research design.	Yes. Participants blinded to study aims.	Yes. Three methods of data collection to triangulate results.	Yes
Recruitment strategy appropriate?	Can't tell. No information provided.	Yes. Participants identified by national coaching director of orienteering	Can't tell. No information provided.	Can't tell. No information provided.
Qualitative Data collection appropriate?	Yes.	Borderline. Administration of structured interview protocol: narrative not provided.	Yes. Informal observations of researchers.	Yes.
Relationship between researcher & participants considered?	No. No critical examination of researcher's own role	Yes.	No. No critical examination of researcher's own role	No. No critical examination of researchers role.
Ethical issues considered?	Can't tell.	Can't tell. No information provided.	Yes. Written consent obtained.	Can't tell. No information provided.
Rigorous data analysis?	Yes, Transcription and *a priori* coding of CRD commentary with thematic analysis.	Borderline. Analysis of SSIs not described.	In terms of qualitative component, yes.	Yes. Transcription and coding of commentaries. In-depth description of analysis: novel evolutionary technique.
Quantitative component	No tool described for assessment of performance: Interrater reliability was good ($r=0.81$, $p<0.1$) Quantitative count of codes: Chi^2 analysis for differences between groups, p and test values stated.	Orienteering checklist: Wilcoxon rank tests to demonstrate differences in scores between groups: No test value. No specific p values (e.g. $p=<0.01$),	i)MacShapa programme ii)Transcription and coding with count of recollections: results demonstrated graphically. No specific numerical values (only range), no formal statistical analysis.	NA
Clear statement of findings?	Yes. Effective decision making associated with greater emotional self regulation & monitoring.	Yes. CRD superior in terms of eliciting thought processes.	Yes. CRD and ACCW both effective. ACCW more time efficient.	Yes. CRD provided insight into cognitive processes. Development of a cognitive model
Value of research	Yes. Implications for training firefighters.	Yes. Discusses future applications in sports research and healthcare.	Implications for how we use information and design of human interface systems.	Yes. Findings will go on to inform the design of a more user friendly cockpit interface for pilots.

Table 2: Quality Analysis. The CASP tool was used to assess quality of the papers included. For mixed methods papers, an extra column addressed quantitative aspects of the research. Colour coding was applied as a quality indicator: Green=no concers; Yellow=minor concerns; Red=major concerns.

RESULTS

Nine studies were included in the synthesis, the features of which are outlined in table 3. There was significant heterogeneity in terms of the participants, which included clinicians (occupational therapists[3,4,5] and emergency physicians[6]) and others (orienteers[7], computer war games players[8], fire station officers[9,10] and military pilots[11]).

There were major differences in how an 'expert' was defined. One study stipulated satisfaction of a six point checklist in order to qualify, which included having been nominated by a hierarchical superior and holding the highest clinical hospital grade[6]. The checklist did not include years of experience, which is known to be a poor surrogate marker for expert performance[12]. Despite this, many papers used this to define their expert population[3,4,5,11]. One study included no definition of the expert[8].

There were small variations in the way in which the intervention was described. True to the original descriptiom[1], all participants wore a head-mounted camera whilst undertaking a complex task and subsequently undertook a facilitated video stimulated 'cued-recall-debrief' process to generate a commentary of the episode. The majority of studies used the authentic 'non-directive' approach to the debrief, however one used a novel, semi-structured format[6]. In five of the nine studies, the individual facilitating the debrief was poorly defined, referred to as "an experimenter or member of the team" and there was no description in any of the studies relating to their level of experience or training.

All studies were observational and had a qualitative component. Four were purely qualitative[3,4,6,11] and the remainder used a mixed methods approach[5,7-10]. In the majority of the studies, the qualitative component involved transcription and coding of the cued-recall-debrief commentary[4-6,8-11]. Two coding approaches were described: a priori or emergent. Generally, where a relevant cognitive model existed in the literature, coding proceeded in an a priori manner[4,5,8-10]. Sometimes, this generated new emergent codes which did not fit with the existing framework[4]. Other studies adopted a wholly emergent code approach to coding[6,11]. Irrespective of how they were generated, emergent codes were used to perform a thematic (or similar) analysis and gain new understanding of cognitive processes or generate novel theoretical models[4,6,11].

Three methods of quantitative analysis were described in the literature:

- Counting the frequency of codes encountered[5,8-10]
- Using a computer programme to generate quantity of codes as a function of time[8].
- A questionnaire measuring participant perceptions of the utility of the intervention in terms of immersion and ability to trigger recollections and insights[7].

Quantitative methods were employed when some form of comparison was taking place;

- Comparing cognitive activity of different participant groups[5] (e.g. novice and expert)

- Comparing level of immersion and recollections generated from cued-recall-debrief with other techniques[7,8] (e.g. free recall)
- Comparing the content of recollections from cued-recall-debrief with other established cognitive models[4,9]
- Stratifying the level of performance through the use of rating scales[10]

The studies could be categorised into three major groups depending upon their aims:

- Basic description of the technique with outcome dats[3].
- Comparing the effectiveness of cued-recall-debrief with other techniques in studying naturalistic decision making[7,8].
- Applications of the information generated from cued-recall-debrief:
    - generation of novel insight or theory into cognitive processes[4,6,9,11].
    - comparing the cognitive processes of novices and experts or good and poor performers[5,10].

Each of these categories will be considered in turn.

Basic descriptive study

A case report by Unsworth demonstrated that occupational therapists and their clients felt that the use of head-mounted camera during their interaction was acceptable[3]. The equipment was portable, did not affect therapist mobility, nor interfere with the therapist-client interaction: participants "forgot" about the camera's presence and were able to "continue as usual". Unsworth was able to generate accurate 'own point-of-view' footage, which displayed the rich visual cues of the client's facial expressions (known to be important in generating immersion). No comment was made as to the extent of experiential immersion experienced by the therapist nor the extent of the recollections generated[3].

The main criticism of Unsworth's paper is that the outcomes were poorly described and interlaced with anecdotal findings from other papers, making identification of novel data challenging.

Comparison with other techniques

Two studies compared the effectiveness of cued-recall-debrief with other methods of studying naturalistic decision making (free recall and a 'walkthrough' technique)[7,8]. Effectiveness was measured in terms of extent of experiential immersion and verbalised recollections .
Omodei et al compared cued-recall-debrief with free recall, in a group of orienteers[7]. Cued-recall-debrief was associated with increased experiential immersion [4.4 vs 6.33 $p<0.01$.] and allowed the individual to recount between two and four times the amount of detail compared to free recall. Cued-recall-debrief generated more recollections related to thoughts [5.3 vs 6.87

p<0.01] and feelings [4.42 vs 6.68 p<0.01]. All participants stated that cued-recall-debrief was the superior method in terms of recalling mental processes.

Elliot et al compared the effectiveness of cued-recall-debrief with a cognitive walkthrough model in participants undertaking a simulated air defence task[8]. The walkthrough model (adversarial crew cognitive walkthrough or ACCW) involved pausing game play when a significant episode occurred in order for researchers to pose questions about the expert's thoughts and feelings.
The paper concluded that ACCW was more time efficient. ACCW took approximately half the time of cued-recall-debrief in generating the same quantity of recollections (1 hour 42 mins versus 3 hours 28 mins).

Compared with ACCW, cued-recall-debrief generated between two and three times the number of recollections related to the expert's own thoughts. The techniques were comparable in their ability to generate recollections relating to their adversary. These results were demonstrated graphically with no specific numerical values for each type of recollection. Furthermore, no formal statistical analysis was undertaken to determine whether the difference between the groups was statistically significant. Experts stated that both techniques were useful and provided rich insights into their cognitive processes. Interestingly, the authors concluded that ACCW was the more effective technique, presumably because it was more time efficient[8].

Applications of cued-recall-debrief

These studies did not evaluate the feasibility or effectiveness of cued-recall-debrief as a means of studying naturalistic decision making; these qualities were assumed. These studies had three distinct aims:

- Generation of additional insight or novel theory into the cognitive processes of the expert[4,6,9,11]
- Comparing the cognitive processes of novices and experts[5]
- Comparing cognitive processes in good versus poor performance[10]

Generation of new theory

Unsworth used two established clinical reasoning models[13,14] as a priori frameworks in order to map the clinical reasoning of occupational therapists[4]. She was not only able to demonstrate the most commonly used types of reasoning (procedural, interactive and conditional) and the way in which they were used (in rapid succession and simultaneously), but she also identified an emergent subcategory of reasoning (generalization reasoning).

Solodilova et al uncovered the cognitive processes of military pilots[11]. Through an inductive, evolutionary data analysis technique, they generated an eight stage model of how pilots acquire and use information, which could be used to help inform the design of more efficient cockpit interfaces. The authors found that pilots used previous experience to generate a model outlining the expected course of the flight which was articulated in a pre-flight brief. Key stages in the flight were bookmarked with expected values, which were compared to actual values. Pilots gathered information from many sources, at two minute intervals and values were always interpreted relative to one another. Information gathering was supported by common information structures such as checklists or standard operating procedures. Upon completion of information gathering, pilots anticipated potential problems. A process of problem solving and information organisation followed before pilots reached a 'point of clarity' and executed their decision. This entire process was converted to knowledge and experience, banked for future flights.

Solodilova's paper gives us some general insights into expert decision making processes. Firstly, experts rely on previous experience to pattern match situations. Secondly, they demonstrate good situational awareness: systematically gathering and assimilating data, whilst anticipating future events. Lastly, they use strategies to reduce the burden on their working memory. Doing things in the same way every time, e.g. using checklists, helps to reduce the burden on working memory and ensures data gathering is systematic[11].

Generation of additional insight

Omodei et al studied the decision making processes of firefighters using cued-recall-debrief[9]. The information gathered by her team was compared to an established decision making model (Recognition Primed Decision Making[15]). Some similarities were demonstrated between the data gathered by CRD and the RPD model (decision making is primarily intuitive), however the former revealed novel information: 20% of cognitive activity was devoted to affect and motivation. This led the researchers to postulate that individuals who are more effective in terms of self-monitoring and emotional regulation may be more skilled performers. This hypothesis was confirmed in their follow-on paper[10], which will be considered later.

Pellacia et al studied emergency department physicians during a patient encounter[6]. The following insights into the decision making processes of participants were revealed:

- All generated at least 4 hypotheses
- Most hypotheses were generated fast and without conscious effort
- Physicians remained vigilant to the emergence of new hypotheses (i.e. avoided premature closure)
- Hypotheses were rank ordered and the order was not consistent with standard probability theory
- Comparison of novice and expert performance

Unsworth et al demonstrated important differences in the clinical reasoning of expert and novice occupational therapists[5]. Whilst experts demonstrated a greater capacity to reason interactively and conditionally with their clients, novices tended to demonstrate a more procedural, textbook approach. In addition, the study showed that experts were more likely to:

- use pattern recognition techniques
- 'think on their feet' (improvisation)

- use more than one type of reasoning at once (multitasking)

Comparing good versus poor performance

Omodei's follow-on study involving firefighters investigated the differences in cognitive activity associated with 'good' versus 'less effective' decision making[10]. All participants were experts and their performance was stratified into poor or good by two senior fire officers using a ten point scale. The performance standard was peer referenced: the six best scores were deemed to represent good performance and the six lowest to equate with poor performance.

There was a significant difference in the way in which cognitive activity was utilised. Poorer performers tended to report a greater frequency of thoughts related to self-monitoring and emotional regulation [X^2 (5;n=485) = 27.4, p<0.001], the content of which was uniformly negative. Good performers thought less frequently about self-monitoring and regulation, but when they did, they exhibited no such negative evaluations. The authors surmised then that effective decision making is associated with fewer negative self evaluations, indicating more effective emotional regulation[10].

The qualitative analysis revealed that the poorer decision makers had a tendency to become overwhelmed by the scenario (cognitive overload). Consequently, the capacity of their working memory was reduced. They demonstrated poorer situational awareness and a tendency to commit decision making pitfalls (e.g. anchoring). They tended to focus on only one salient feature of the situation whilst ignoring emerging threats[10].

More skilled participants demonstrated better situation awareness as they:

- kept track of events
- anticipated developments

Good performers also reacted to new threats without undue irritation or concern. Interestingly, there was no correlation between level of performance and number of years of experience[10].

DISCUSSION

In summary, the literature shows that video stimulated cued-recall-debrief is a useful and feasible means of gaining insight into expert performance. The head-mounted camera is portable and does not appear to interfere with the episode being studied (though this latter observation is purely anecdotal)[3]. Cued-recall-debrief (CRD) is associated with a high level of experiential immersion and allows an individual to recount between two and four times the number of recollections when compared with free recall[7]. The method is particularly useful in terms of gaining insight into one's own actions and feelings[7,8]. It is however more time intensive in terms of time than other information elicitation techniques such as cognitive walkthrough[8]. Clearly, a cognitive walkthrough is not always possible in naturistic decision making settings. For example, it would be dangerous to ask a surgeon to pause and discuss their thoughts and feelings when performing an emergency, tracheostomy.

The majority of the papers included in this review considered the technique's applications:

- Generating additional insight or novel theory relating to expert cognition
- Comparing the cognitive processes of two groups of individuals undertaking the same task

These studies demonstrate commonalities in the cognitive and affective processes of experts across several disciplines. Generally, experts tended to use intuitive, as opposed to analytical, reasoning. These types of reasoning can be conceptualised using the dual process theory which posits that two systems of decision making exist[16]. System 1 is intuitive, automatic and fast and largely based upon pattern matching. It relies upon having built up a bank of previous experience. System 2 is slower and analytical and involves a degree of problem solving.

Other features demonstrated by experts were as follows:

- An ability to multitask
- An ability to improvise
- Awareness of and ability to use strategies to avoid common pitfalls: checklists, cognitive strategies (avoiding premature closure)
- Good situational awareness.
- Effective self-monitoring and emotional regulation

The final point relating to emotional regulation is in keeping with what we know from some of the existing decision making literature. Too much concern about how well one is doing in a task sometimes disrupts performance by loading short term memory with pointless thoughts. As controlling thoughts and behaviours is one of the tasks performed by system 2, perhaps experts have a more efficient or better developed sense of analytical reasoning[17].

A common limitation of the studies was a lack of transparency as to the relationship of the lead researcher to the activity being studied. As a researcher, being both an 'insider' and 'outsider' can have an impact upon the way in which data is collected and analysed and therefore it is important to reflect and comment on this influence[18].

One paper applied no formal statistical analysis when comparing numerical values between two groups[8]. This makes it difficult for us to surmise whether the differences between the groups were statistically significant.

For the majority of the studies, the debrief facilitator and their level of experience was poorly defined. Facilitating a debrief is a difficult skill that requires practice. As a beginner, there is a natural tendency to ask 'why', leading the expert to evaluate their performance, rather than to facilitate the recall of their cognitive and affective processes. To maximise the integrity and reproducibility of the technique, we should ensure that reporting is explicit in terms of the debriefer's training experience and introduce a recommendation for core debriefing competencies to be met ahead of engaging with the technique.

CONCLUSIONS

To the authors' knowledge, this is the first systematic review considering the current evidence base for video stimulated cued-recall-debrief. The authors tentatively conclude that the technique appears to be a feasible and valid method of gaining insight into an expert's thoughts and feelings. Additionally, it can be a useful tool in building models and patterns of cognitive activity and characterising the way in which more and less skilled individuals tend to think and feel. These results should however be interpreted with caution given the methodological flaws described. Future research might include expanding the use of the technique in medicine; with attention to technical and non-technical skills. Specifically, this could examine:

- Whether there are particular environments, procedures or individuals for which this technique works best.
- Whether the insights gained into the cognitive processes of experts are transferrable to a group of learners.
- The impact of the intervention on the activity being studied. Although inanimate technologies and materials tend to be accepted as part of the backdrop in medicine, a head-mounted camera may influence what people think, say and do.

REFERENCES

1. Omodei M, Wearing A, McLennan J. Head-mounted video recording: A methodology for studying naturalistic decision making, in: *Decision Making Under Stress: emerging themes and applications*. London: Ashgate; 1997.

2. *Critical Appraisal Skills Programme.* Available at https://casp-uk.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-Download.pdf [Date of access: July 2017]

3. Unsworth CA. Using a head mounted camera to study clinical reasoning. *American Journal Occupational Therapy* 2001; 55, 582-588.

4. Unsworth, C. A. Using a head-mounted video camera to explore current conceptualizations of clinical reasoning in occupational therapy. *American Journal Occupational Therapy*. 2005; 59, 31–40.

5. Unsworth CA. The Clinical Reasoning of Novice and Expert Occupational Therapists. *Scandinavian Journal Occupational Therapy* 2001; 8(4):163-173

6. Pelaccia, T, Tardif J, Triby E, Ammirati C, et al. How and When Do Expert Emergency Physicians Generate and Evaluate Diagnostic Hypotheses? A Qualitative Study Using Head-Mounted Video Cued-Recall Interviews. *Annals of Emergency Med* 2014; 64(6); 575-585.

7. Omodei M, McLennan J. Studying complex decision making in natural settings: Using a head mounted camera to study competitive orienteering. *Perceptual and Motor skills* 1994; 79; 1411-1425.

8. Elliott G, Crawford J, Solodilova I (aka Whiteley I), Sanderson P. Comparison of Adversarial Crew Cognitive Walkthrough vs Cued Debrief Recall Techniques for Eliciting Information in a D3M environment. Proceedings of the 2000 Annual Conference of the Computer-Human Interaction Special Interest Group of the Ergonomics Society of Australia. Sydney, Australia.

9. Omodei M, McLennan J. How expertise is applied in real world dynamic environments: Head-Mounted Video and Cued recall as a methodology for studying routines of decision making; Study 1, in: *The Routines of Decision Making*. New Jersey: Lawrence Erlbaum Associates; 2004.

10. Omodei M, McLennan J. How expertise is applied in real world dynamic environments: Head-Mounted Video and Cued recall as a methodology for studying routines of decision making; Study 2, in: *The Routines of Decision Making*. New Jersey: Lawrence Erlbaum Associates; 2004.

11. Solodilova-Whiteley I, Johnson P. Uncovering the information needs in complex aerospace systems. *Journal Reliability Engineering & System Safety* 2006; 1566-1575.

12. Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Annals Internal Medicine* 2005; 142:260-273.

13. Mattingly, C., & Fleming, M. H. *Clinical reasoning: Forms of inquiry in a therapeutic practice.* Philadelphia: F. A. Davis; 1994

14. Schell, B. A., & Cervero, R. M. Clinical reasoning in occupational therapy: An integrative review. *American Journal of Occupational Therapy* 1993; 47; 605–610.

15. Klein G. The recognition-primed decision (RPD) model: Looking back, looking forward in Naturalistic decision making, in: Expertise: *Research and applications. Naturalistic decision making*. Hillsdale, US: Lawrence Erlbaum Associates; 1997.

16. Kahneman, D., & Frederick, S. Representativeness revisited: Attribute substitution in intuitive judgment, in: *Heuristics and biases: The psychology of intuitive judgment*. New York, US: Cambridge University Press; 2002.

17. Kahneman, D. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux; 2011.

18. Dwyer S, Buckle J. The Space Between: On Being an Insider-Outsider in Qualitative Research. *International Journal of Qualitative Methods* 2009. 8(1)

Figure 1: PRISMA diagram (Preferred Reporting Items for Systematic Reviews and Meta-analyses).

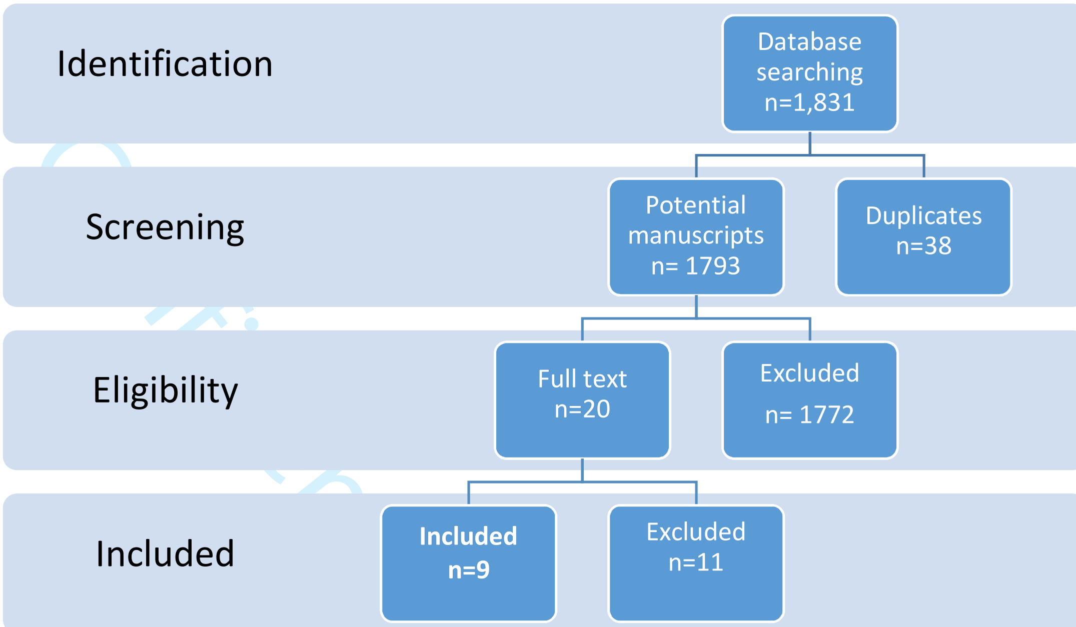

Table 3. Features of the studies included.

Paper	Omodei	Elliot	Unsworth	Unsworth
Publication year	1994	2000	2001	2001
No. participants	6	4	1	3 (+2 novice)
Participant type	Orienteers	Computer war games players	Occupational therapists (OTs)	Occupational therapists
Gender	3M: 3F	All male	All female	All female
Mean age	16	NA	NA	NA
Design	Crossover. Mixed methods.	Between subjects comparison. Mixed methods.	Case report.	Mixed methods
Aim	Compare two methods for studying naturalistic decision making (NDM)	Compare two methods of studying NDM	Basic description of technique	Compare cognitions of experts and novices
Context	Competitive orienteering circuits	Mission commander in simulated air defence tasks	An encounter between OT and client during a physical rehabilitation episode	An encounter between OT and client during a physical rehabilitation episode
Cognitions relate to	Navigation related thoughts and feelings	Situation, intentions and constraints in relation to self and adversary	Clinical reasoning	Clinical reasoning
Comparator	Free recall	Adverserial crew cognitive walkthrough (ACCW)	-Concurrent reporting -Free recall -Retrospective reporting with assisted recall (audio or external video)	No
Primary outcome	-Level of experiential immersion -Extent of insight into decision making processes	- Extent of insight into decision making processes	-Feasibility -Level experiential immersion (surrogate markers of POV camera and visibility of cues)	NA
Methods of data collection and/or analysis	1. Orienteering Review Analysis checklist (quantitative) 2. Structured interview protocol comparing each technique (qualitative)	1. MacShapa programme: quantity of information taken as a function of time (quantitative). 2. Transcription and coding of commentaries with quantitative count of recollections. 3. Informal researcher observations (qualitative).	Author's own experiences	NA
Secondary outcome	NA	NA	NA	Compare the cognitions of experts versus novices
Method of data collection and/or analysis	NA	NA	NA	Transcription and coding of commentaries. Quantitative count of recollections and qualitative thematic analysis
Coding approach	NA	*A priori* – framework not stated	NA	*A priori* (Mattingly and Fleming, Schell and Cervero clinical reasoning frameworks)
Statistical method (quantitative only)	Wilcoxon signed rank test comparing differences between techniques	Basic descriptive, no formal statistical analysis applied	NA	Chi squared analysis for differences between groups. Simple descriptive statistics.
Potential applications CRD	Training students/beginners to think like experts	Training students/beginners to think like experts	Training	Training. Trainers can target the cognitive gap that separates novices and experts

Paper	Unsworth	Omodei	Omodei	Solodilova	Pellacia
Publication year	2005	2005	2005	2006	2014
No. participants	13	9	12	4	15
Participant type	Occupational therapists	Fire station officers	Fire station officers	Military pilots	Emergency physicians (EPs)
Gender	All female	All male	All male	All male	11 males, 4 females
Mean age	NA	NA	NA	NA	42
Design	Qualitative	Mixed methods	Mixed methods	Qualitative	Qualitative
Aim	Conceptualise the clinical reasoning of OTs	Conceptualise the decision making of firefighters	Compare the cognitions of good versus poor decision makers	Conceptualise information use by pilots	Conceptualise the clinical reasoning of EPs
Context	Client encounter during physical rehabilitation	Simulated emergency fire incident	Simulated emergency fire incident	Simulated flight	Encounter with patient with life threatening condition
Recollections relate to	Clinical reasoning	Situation assessment, intention formation, self-monitoring and regulation	Situation assessment, intention formation, self-monitoring and regulation	Gathering and assimilation of information, anticipating future states	Clinical diagnosis
Comparator	No	No	No	No	No
Primary outcome	No	No	No	No	No
Methods data collection/analysis	NA	NA	NA	NA	NA
Secondary outcome – application	Generate additional insight into clinical reasoning of OTs	Generate additional insight into how cognitive activity is utilised in firefighters	Evaluate differences in cognitive content between poor and good decision makers	Generate a model demonstrating information gathering and assimilation in a cockpit.	Define the cognitive strategies of ED physicians in making a clinical diagnosis
Method of data collection and/or analysis	Transcription and coding of CRD commentary.	Transcription and coding of CRD commentary with quantitative count of recollection types. *Pattern compared to Klein's Recognition Primed Decision Model.*	Transcription and coding CRD commentary. Quantitative count of type of recollections and qualitative thematic analysis. 10 point rating scale to rate and stratify performance.	Transcription and coding of CRD commentary in order to generate a model of information flow.	Transcription and coding of CRD commentary with thematic analysis
Coding approach (framework used)	*A priori* (Mattingly and Fleming, Schell and Cervero clinical reasoning frameworks)	*A priori* (Cognitive Process Tracing Categorisation Scheme (CPTCS))	*A priori* (CPTCS)	Emergent	Emergent
Statistical method (quantitative only)	NA	NA	Chi^2 analysis comparing differences between groups.	NA	NA
Utility	Training students/beginners to think like experts	Training	Training. Trainers can target the cognitive gap that separates novices and experts	Implications for human/ systems interface design.	Training: Trainers can target the cognitive gap that separates novices and experts.