

Macquarie University ResearchOnline

This is the author version of an article published as:

Loveday, Thomas; Wiggins, Mark W.; Harris, Jemma M.; O'Hare, David and Smith, Neil (2013) An Objective approach to identifying diagnostic expertise among power system controllers, *Human Factors*, Vol. 55, No. 1, p.90-107

Access to the published version: <http://dx.doi.org/10.1177/0018720812450911>

Copyright: Copyright the Authors 2013. Version archived for private and non-commercial use with the permission of the author/s and according to publisher conditions. For further rights please contact the publisher.

Topic: Cognitive Processes

An Objective Approach to Identifying Diagnostic Expertise Amongst Power System
Controllers

Loveday, T., Wiggins, M.W., Harris, J.M.
Macquarie University, Sydney, Australia

Smith, N.
TransGrid, Sydney, Australia

O'Hare, D.
University of Otago, Dunedin, New Zealand

Author Note

Thomas Loveday, Department of Psychology, Macquarie University; Mark Wiggins,
Department of Psychology, Macquarie University; Jemma Harris, Department of Psychology,
Macquarie University; David O'Hare, Department of Psychology, University of Otago; Neil
Smith, TransGrid.

This research was supported in part by grants from the 'Australian Research Council'
and TransGrid Pty Ltd under the former's Linkage Program (Grant Number LP0884006).

Correspondence concerning this article should be addressed to Thomas Loveday,
Department of Psychology, Macquarie University, NSW, Australia, 2109.

E-mail: thomas.loveday@mq.edu.au

Keywords: cues, diagnostic performance, competent, expert, assessment

Word count: 6381

Abstract

Background: Despite its increasing importance in the contemporary workplace, there is little understanding of the cognitive processes that distinguish novice, competent and expert performance in the context of remote diagnosis. However, recent evidence suggests that cue acquisition and utilization may represent a mechanism by which the transition from novice to expertise occurs.

Objective: The present study investigated whether performance across a range of cue-based cognitive tasks differentiated the diagnostic performance of power control operators into three distinct groups, characteristic of novice, competence, and expertise.

Method: The study involved the application of four distinct cue-based tasks within the context of power system control. Sixty-five controllers, encompassing a range of industry experience, completed the tasks as part of an in-service training program.

Results: Using a cluster analysis, it was possible to extract three distinct groups of operators on the basis of their performance in the cue-based tasks, and these groups corresponded to differences in diagnostic performance.

Conclusion: The results indicate assessments of the capacity to extract and utilize cues were able to distinguish expert from competent practitioners in the context of power control.

Application: Assessments of the capacity to extract and utilize cues may be used in the future to distinguish expert from non-expert practitioners, particularly in the context of remote diagnosis.

An Objective Approach to Identifying Diagnostic Expertise Amongst Power System Controllers

In many complex industrial environments, including aviation and power system control, humans now share the responsibility for the safety and security of the system with advanced technology (Flaspöler et al., 2009). This interdependence between human operators and technology has generally resulted in improvements in both the efficiency and the accuracy of system performance. However, it has also shifted the role of the operator from ‘active controller’ to ‘system monitor’, with a resulting emphasis on diagnostic reasoning.

Diagnostic reasoning is a cognitive skill whereby patterns of information arising from the environment are matched to pre-existing patterns of information resident in long-term memory, which are then used to form an explanation of a previous event and/or to anticipate or predict a future event (Schimdt & Boshuizen, 1993). While successful diagnosis is possible with limited knowledge and instruction (O’Hare, Mullen, Wiggins, & Molesworth, 2008), skilled diagnostic performance is typically acquired through extensive experience and practice, and is generally regarded as the province of domain experts (Gale & Walker, 1993; Jarodzka, Scheiter, Gerjets, & van Gog, 2010; Wallis & Horswill, 2007; Wiggins & Bollwerk, 2006; Wiggins & O’Hare, 1995). Therefore, understanding the cognitive processes that enable the acquisition of expert diagnostic skills has important implications for workplace assessment, evaluation and training.

The acquisition of cognitive skills such as diagnostic reasoning has typically been investigated using comparative analyses of experts and novices. Referred to as the expert-novice paradigm, it has been a useful strategy to identify the features of cognitive processing that differentiate expert from novice performance. Moreover, the fact that these comparisons have been undertaken across many different domains enables the identification of similarities that reflect broader underlying constructs.

One of the most robust outcomes of the expert-novice paradigm is that experts can maintain and recall large amounts of information in short-term memory (de Groot, 1965; Schimdt & Boshuizen, 1993) despite the well-established processing restriction of seven simultaneous chunks (Miller, 1956). This suggests that experts have developed particular cognitive strategies that enable them to acquire and integrate domain-related information quickly and accurately.

Ericsson and Kintsch (1995) suggested that the accuracy and the capacity of memory-based recall amongst experts reflects the utilization of cues to selectively attend to, and integrate, task-related information. Similarly, Klein (1989) described the utilization of cues as one of the distinguishing characteristics of domain experts. These cues are acquired through their association with specific events, eventually becoming automated with repeated exposure (Lipshitz, Klein, Orasanu, & Salas, 2001; Wiggins & O'Hare, 2003). Therefore, the application of cues effectively provides a mechanism to reduce the cognitive demands of a task, without the loss of relevant diagnostic information (Rasmussen, 1983).

Empirical support for the application of cues in expert diagnosis is reasonably well established (Cellier, Eyrolle, & Marine, 1997; Lipshitz, et al., 2001; Schriver, Morrow, Wickens, & Talleur, 2008). For example, in cricket, expert batsmen use movement cues to predict an opposing bowler's intentions, and to prepare their shots before the ball has been delivered (Müller, Abernethy, & Farrow, 2006). Similarly, in the aviation context, there is evidence to suggest that expert pilots focus on those cues that are most predictive of engine failure, and similarly, those cues that are most likely to indicate the source of a fault once an engine failure has occurred. Accordingly, expert pilots respond faster and more accurately to faults than novice pilots (Schriver, et al., 2008).

Although the proposition that experts use cues is not in dispute, a major limitation of previous research based on the expert-novice paradigm is that experts have normally been

identified simply on the basis of their cumulative experience operating within a particular domain. For example, Schraagen (1993) defined psychology research design experts as those individuals with at least ten years of experience designing experiments in psychology. Similarly, Coderre, Mandin, Harasym and Fick (2003) defined gastroenterology experts as qualified specialists with more than five years of experience. However, several studies have demonstrated that individuals with extensive experience do not necessarily exhibit superior performance to novices on domain-representative tasks (Ericsson & Lehmann, 1996; Shanteau & Stewart, 1992). It is possible that this may reflect the fact that 'years of experience' in a role does not necessarily equate directly to the progression towards expertise. Therefore, experience may be a necessary, but not sufficient, precursor to expertise. Dedicated practice is also required, and this is likely to be best reflected in performance-based measures (Charness, 1976; Ericsson & Charness, 1994; Ericsson & Ward, 2007).

A second, but related, limitation of expert-novice comparisons is that this approach may not capture the full spectrum of cognitive skills development. Expert-novice comparisons, by definition, can only capture the two extremes of performance. However, amongst skill acquisition theorists, there is broad consensus that at least three stages comprise the progression from novice through competence, to expertise (Wiggins & O'Hare, 1993). For example, Ericsson and Lehmann (1996) refer to three stages of cognitive skill acquisition, each of which is associated with a distinct level of performance. These include a novice phase, where the component skills and knowledge are developed, a competent phase, in which learned skills have become partially or entirely automated through associations in memory, through to the expert phase, characterized by the ability to maintain control over the strategic aspects of performance, facilitating the selective encoding and refinement of those associations. Similarly, Anderson (1983) proposed a three-stage model of skill acquisition,

which broadly maps to novice, competent and expert phases described by Ericsson and Lehmann (1996).

The difficulty associated with stage models of skill acquisition is that there is an assumption that the acquisition of particular skills within one stage necessarily facilitates the progression to subsequent stages of skill acquisition. In fact, there is evidence to suggest that many operators can fail to progress beyond competence, despite the accumulation of significant amounts of operational experience and practice (Doane, Pellegrino, & Klatzky, 1990; Reif & Allen, 1992; Watson, Stimpson, Topping, & Porock, 2002). Therefore, it is necessary to compare such individuals to novice and expert practitioners to improve our understanding of the progression towards expertise. However, to ensure the reliability and validity of such comparisons, a priori techniques for distinguishing genuinely competent individuals from novices and experts must be devised.

The Proposed Solution

Although many researchers have used indirect measures of expertise, such as operational experience, as the basis for classifications, the identification of genuine experts using domain-related performance is well established. For example, both deGroot (1965) and Elo (1986) identified chess experts on the basis of tournament rankings, while Ericsson (2004; Ericsson, Charness, Feltovich, & Hoffman, 2006) identified sports, medical, research-design and music experts using domain-representative tasks. Similarly, several researchers have stratified task performance using a within-task criterion. For example, movement responses (Vaeyens, Lenoir, Williams, & Philippaerts, 2007) and visual search (Savelsbergh, Van der Kamp, Williams, Ward, & Brogan, 2005) techniques in laboratory environments have been used to predict soccer performance in the field.

The present study extended this work by proposing that it may be possible to stratify diagnostic reasoning performance, utilizing tasks that capture distinct aspects of broader

domain-related performance. To capture the three-stages of cognitive skill development, whereby novices' progress to expertise based on the acquisition and refinement of cue-associations in memory (Klein, 1989), tasks were selected on the basis that the judicious utilization of cues would facilitate performance. Four tasks were developed for the present study, comprising feature identification, feature discrimination, background knowledge, and information acquisition. Each task was modified to utilize stimuli from the power system control domain. Power system control was selected as the domain of interest on the basis that the activity is largely cognitive in nature and, as a computer-based activity, it was possible to ensure a degree of experimental control while maintaining a high degree of ecological validity. Further, the domain is one in which operators develop highly specialized skills, and thus, have the opportunity to develop expertise.

Feature identification is based on the observation that experts utilize a limited set of features that either: (a) provide a degree of ecological validity and/or (b) are particularly consequential to the safety and security of the system (Müller, et al., 2006; Schriver, et al., 2008) The identification of these key features restricts the search for information and thereby primes a response so that response latency is decreased, yet remains accurate (Schyns, 1998). The skill of feature identification can be examined through a target-matching task in which the operator is asked to identify a domain-relevant target from an array. Because power-controllers must frequently extract network fault alerts from complex displays, a process critical to the initial recognition of network problems, the present study used a network fault alert within a power network display as the target (see Figure 1). This task can also be extended, by asking the operator to identify, with only limited exposure, the location of a particular target. In both cases, this requires the capacity to identify key features from a complex scene, a skill that is presumably most developed amongst experts (Ratcliff & McKoon, 1995; Santos & Badre, 1994) and is a known mediator of diagnostic performance

(Strauss & Kirlik, 2006). Thus, domain experts can be expected to demonstrate reduced latency and greater accuracy in response to the various cue-based tasks (Schriver, et al., 2008; Schyns, 1998).

Figure 1 About Here

Feature discrimination is broadly based on the Lens Model's conceptualization of expertise, whereby performance can be facilitated by the consistent recognition and utilization of optimally diagnostic cues (Brunswik, 1956). Consistent with this view, Weiss and Shanteau (2003) demonstrated that experts' perception of the relative utility of features remains consistent over a series of trials. Summarized by the Cochran-Weiss-Shanteau (CWS) index of expertise, Weiss and Shanteau (2003) proposed that the ratio of discrimination over inconsistency could be used as a metric of expertise. *Discrimination* refers to the differential evaluations of the perceived utility of individual features within a set, whereas *consistency* refers to the participant's evaluation of the same feature over time. Thus, the ratio will be larger when participants both discriminate effectively and are consistent in their ratings of utility for the same features.

While the CWS index is undoubtedly a useful tool for the assessment of expertise, its validity relies on the use of a series of simple scenarios that are totally consistent in terms of the available cues. However, the nature of many complex industrial environments is such that the development of such scenarios requires that they be simplified to the point where key features associated with the actual task may inadvertently be removed, or alternatively, become so obvious as to artificially reduce the variation in responses between participants. Consequently, the researcher may impose structure and rules on judgments that may not actually exist in the naturalistic setting (Kahneman & Klein, 2009).

The present study took an alternative approach, by analyzing responses to a single, complex scenario for which the relevance of a range of features was considered. In the field, it is critical that power-controllers are able to selectively ignore irrelevant alerts and information and focus on those indicative of the causal fault. Therefore, consistent with the principles underlying the CWS index, it was argued that, in comparison to non-experts, experts would be better able to discriminate between features with relatively greater or lesser relevance given a specific scenario. Based on the discrimination component of the CWS formula, discrimination was calculated using within-subject response variance in ratings of utility across each of the features (Weiss & Shanteau, 2003). Thus, experts were expected to demonstrate relatively greater variance in their ratings across all features, regardless of utility, in comparison to non-experts (Kolodner, 1997; Weiss & Shanteau, 2003).

Although paired associations are not closely matched to a real-world process, this measure is based on the long-standing practice of measuring the strength of associations in memory by presenting word pairs and measuring response latency (Ackerman & Rathburn, 1984) or ratings of perceived relatedness (Schvaneveldt, Beringer, & Lamonica, 2001). Therefore, this task should provide a measure of the real-world ability to recognize the relationship between key features and events, critical to the recognition of latent faults. In its present form, the task presents two random, but domain-relevant words or phrases either sequentially or simultaneously, and measures response latency and variance in ratings of relatedness. The sequential pairings were included to assess the impact of priming on response latency (Klein, 1989), whereas the simultaneous pairings were included to assess the participant's ability to divide attention (Sweller, 1988).

Amongst experts, cue associations are refined and efficient, so that the presentation of the feature within a valid cue should prime the outcome and vice versa, resulting in lower response latencies. This is consistent with the paired-relatedness analyses utilized by

Schvaneveldt, Beringer and Lamonica (2001) whereby expert aviators demonstrated reduced response times and greater discrimination for ratings of association between paired items. As with the Feature Discrimination Task, discrimination in the paired association task was measured using the within-subject variance across item ratings, and thus, domain experts were expected to demonstrate both reduced response latency (Ackerman & Rathburn, 1984; Schvaneveldt, et al., 2001) and greater response variance in the paired association task (Weiss & Shanteau, 2003).

The *Transitions Task* is based on the observation that experts and non-experts differ in the sequence in which they acquire task-related information. Both Wiggins and O'Hare (1995) and Wiggins and Bollwerk (2006) were able to differentiate expert from non-expert pilots based on the sequence in which information was accessed when responding to a failure scenario. Where non-experts tended to acquire information in a sequence consistent with its physical presentation, domain experts tended to acquire information less sequentially, in an order reflecting the relevance and context of the information.

Wiggins and O'Hare (1995) and later Wiggins, Stevens, Howard, Henley, and O'Hare (2002) have developed a procedure to quantify the transition involved in the acquisition of information using the sequence of information screens as they are presented to participants as a reference score. For each participant, sequentially accessed pairs of information screens were examined to establish whether they were drawn from the same or different sub-menus. On the basis of these data, it was possible to establish the proportion of pairs of information screens accessed from the same submenu and thus, the extent to which information was accessed according to the structure in which it was presented.

Although Wiggins and O'Hare (1995) employed a menu structure, the present study established the transition sequence using a single list of domain-relevant categories. Each category on the list could be clicked to reveal additional information regarding a network

incident. The information and structure was consistent with reports provided by the system. By examining the sequential pairs of information screens as they were accessed and establishing whether they represented the sequential pairs of information screens as they were listed, it was possible to calculate the proportion of pairs of information screens accessed in the sequence in which they were listed. Based on Wiggins and O'Hare (1995), experts were expected to utilize a less sequential strategy for acquiring information, and thus, access a lower proportion of screens sequentially.

Aims and Hypothesis

The specific aim of the present study was to determine whether, collectively, four approaches to the measurement of cue utilization – feature identification, feature-discrimination, paired associations, and the transition task – were able to decompose performance into three levels consistent with the novice, competent and expert stages of skill acquisition described by Ericsson and Lehmann (1996). Because experts rely heavily on cue-associations in memory to direct their performance (Klein, 1989), they were expected to demonstrate consistently superior performance on those reasoning tasks that required the judicious extraction and utilization of cues. By comparison, competent non-experts, who although experienced (Gray, 2004), draw upon less refined cases or cue-associations (Anderson, 1983), should demonstrate somewhat weaker performance. Novices, with limited operational experience, would be expected to perform relatively poorly on any specialist task. Therefore it was hypothesized that:

- 1. Performance across the tasks would cluster into three levels.*
- 2. These clusters, reflecting novice-, competent- and expert-practitioners, would differ in terms of performance in each of the tasks. The expert cluster, compared to the other clusters, was expected to be associated with:*

- a. *A reduced response latency and increased accuracy during the feature identification task;*
 - b. *A reduced response latency and an increase in the variance of ratings of association between pairs in the paired association task;*
 - c. *An increase in the variance of ratings of utility of individual features during the feature-discrimination task; and*
 - d. *A decrease in the proportion of information screens accessed sequentially during the transition task.*
3. *Novice and competent clusters would be differentiated on the basis of years of experience, but that no difference would be evident between competent and expert clusters.*

The final hypothesis was designed to test the proposition that the expert cue utilization cluster was associated with diagnostic expertise. Because skilled diagnostic performance is generally regarded as characteristic of domain expertise (Gale & Walker, 1993; Jarodzka, et al., 2010; Wallis & Horswill, 2007; Wiggins & O'Hare, 2006), it was also hypothesized that:

4. *In comparison to the other clusters, the expert cue utilization cluster would select the optimal diagnostic outcome with greater frequency.*

Method

Participants

The participants represented a convenience sample recruited during a training session held by TransGrid, the owner and operator of the New South Wales electrical power transmission network. They comprised 65 power transmission operators, of whom 61 were male and four were female. All of the participants approached elected to participate in the study. They ranged in age from 23 to 61 years with a mean of 38.14 years (SD = 12.63). The

participants had accumulated between 0 and 40 years of experience in power transmission, with a mean of 14 years ($SD = 13.30$). Twenty-seven participants were part of a graduate program, and thus, were considered inexperienced in power transmission, with a mean of 2.85 years of experience ($SD = 1.76$). The remaining 38 participants were not recent graduates, and had accumulated a mean 22.36 years of experience ($SD = 11.98$) within the domain.

Stimuli Development

Prior to the study, cognitive interviews were conducted with a subject matter expert from the power transmission domain to develop the stimuli. The subject matter expert was selected on the basis of peer reference and position. Consistent with the critical incident technique (Flanagan, 1954), he was asked to describe operational incidents in which he had made a critical decision and which had occurred within the preceding six months. It was emphasized that the incidents reported should be complex due to the potential for uncertainty and where he felt that his expertise and experience made a critical difference to the outcome.

The subject matter expert described several incidents in detail, including a comprehensive description of all of the information that might have been available, even if the subject matter expert believed that that information was irrelevant. This approach was taken to circumvent the possibility that such a skilled practitioner may have been unaware of the cues that guided their decision-making (Kahneman & Klein, 2009). The descriptions were corroborated against incident reports and network event history logs recorded by the network control system.

Two incidents were subsequently organized into timelines reflecting, as accurately as possible, the actual sequence and timing of events, and formed the basis of the Feature Discrimination (See *Appendix A: The Feature Discrimination Task*) and Transition Task (See *Appendix B: The Transition Task*) scenarios respectively. The remaining incidents were used

to develop a catalogue of features that an individual may or may not consider during a complex diagnostic task within power control. These features were then paired to form the items in the Paired Association Task (See *Appendix C: Paired Association Task Items*). The images used in the Feature Identification Task were screen captures of actual line failure alerts, recorded by operators as they occurred (See Figure 1).

Stimuli

This study involved the use of EXPERTise (Wiggins, Harris, Loveday, & O'Hare, 2010). EXPERTise is a 'shell' software package designed in-house, to record performance across four, cue-based expert reasoning tasks: the Feature Identification task, the Paired Association task, the Feature Discrimination task, and the Transition task. EXPERTise was specifically designed so that these tasks could be customized to match the stimuli used in a particular domain.

The Feature Identification task incorporated two stages. In the first stage, participants were presented with Supervisory Control and Data Acquisition (SCADA) interface screen captures of the transmission network (see Figure 1). Each SCADA screen displayed the failure of a circuit breaker. The distance from the target and the speed with which participants clicked on the failed circuit breaker was recorded. In the second stage, the participants were shown a SCADA screen capture for 1.5 seconds, during which a failed circuit breaker was displayed. On a subsequent screen, they were asked to select the location of the fault from four possible options (designated by codes associated with different power lines) and this response, together with the response latency, was recorded.

The Paired Association task also incorporated two sub-stages. In the first sub-stage, participants were presented with pairs of features and events sequentially. For example, the participants might be shown the phrases 'Energies' and 'Current'. The stimuli were presented as text and displayed for 1.5 seconds. The participants rated the perceived strength of the

relationship between the features and events on a six-point scale. The second stage also involved the presentation of features and events for 1.5 seconds but the stimuli were presented simultaneously on the screen. The participants' ratings and their response latencies were recorded.

The Feature Discrimination task involved the presentation of a SCADA screen capture that incorporated a system failure. Participants were asked to assess the situation and then, on the same screen, select one of four possible responses to the failure, including a '*do nothing*' option. On the following screen, they rated, on a six-point scale, the utility of each of nine individual features of the scenario in formulating their response. The participants' response to the scenario, and their ratings of each feature, were recorded.

The Transition task involved the presentation of a descriptive scenario that required participants to acquire additional information prior to selecting a response. The scenario involved an unverified report of a child climbing a high-tension power line. The information presentation was intentionally incomplete so that participants would be required to acquire additional information to form an accurate diagnosis and response. The additional information was presented in the form of a list of information screens. Each information screen was accessed through a mouse click on the relevant button and this enabled the process of information acquisition to be traced. Each button was labeled with a relevant term (see *Appendix B: The Transition Task*) to ensure that its contents were clear to participants and this was verified during usability testing.

Procedure

The participants were invited to participate in the study during a work-related training session. Conducted in a group of approximately ten participants, they were briefed on the purpose of the study and then asked to sign a consent form if they wished to continue. They completed a brief demographics questionnaire, which captured the general (years in the

domain) and specific (years in their current position) experience of the operators, and they then began the EXPERTise test battery via laptop computer. On completion of EXPERTise, they answered a second questionnaire, this time regarding their perception of the EXPERTise tasks. The EXPERTise session and questionnaires took approximately 40 minutes to complete.

Results

Descriptive Data

Each continuous variable was screened for normality of distribution. For all of the variables, the skewness and kurtosis figures suggested a non-normal distribution. This was confirmed by visual inspection of the corresponding histograms. The variables were then subjected to a square root transformation. Subsequent analysis revealed that the skewness and kurtosis figures were within acceptable limits for most of the variables, skewness < 1 , kurtosis < 1 . The exception, the reaction time scores for the Feature Identification task, was subjected to a log-transformation, but this too failed to meet acceptable limits: skewness > 1 , kurtosis > 1 . Based on concerns that this skewness indicated that the measure lacked precision or was subject to confounding variables, the data were excluded from further analysis.

Inter-Correlations

The Pearson bivariate correlations between each of the EXPERTise measures are listed in Table 1. If the measures assessed independent components of performance, only weak to moderate correlations should have been observed between each of the variables.

Table 1 About Here

Statistically significant and sizeable correlations were observed between the response latencies for the sequential and simultaneous Paired Association tasks, $r = .93$, $p < .01$, and the variance scores for the same tasks, $r = .94$, $p < .01$. These relationships might be expected since the sequential and simultaneous presentation of feature-event pairs merely represented two approaches to assessing the same aspect of diagnostic expertise.

The response latencies for the sequential and simultaneous forms of the PAT also had a moderate to large correlation with Feature Discrimination Task ratings variance, $r = .49$, $p < .01$, and $r = .52$, $p < .01$, respectively. Similarly, response latencies for the sequential and simultaneous forms of the PAT had a weak to moderate correlation with Feature Identification Accuracy scores, $r = -.26$, $p < .05$, and $r = -.31$, $p < .05$, respectively.

Correlations with Experience

To investigate the relationship between years of domain-general and job-specific experience and each of the EXPERTise measures, bivariate correlations were undertaken between years of experience (domain general and domain specific) and performance on the EXPERTise tasks. Table 2 details the correlations between performance and demographics.

Consistent with the view that experience is a necessary, but not sufficient, requirement of expertise, there was a weak to moderate correlation between years of experience in the domain and the measures of task performance, $.13 < r < .33$.

Table 2 About Here

Cluster Models

The primary aim of the present study was to investigate the feasibility of classifying participants into three groups on the basis of observed levels of performance across the EXPERTise tasks. Due to differences in scale, the variables were standardized prior to

clustering, as is required by the K-means cluster procedure of the SPSS statistical package (Tabachnick & Fidell, 1996). Based on the theoretical rationale that performance can be decomposed into novice, competent and expert levels, three clusters were entered as the most likely fit to the data.

Table 3 presents the results of the cluster analysis, the mean score for each cluster on each of the outcome variables, and the cluster rankings for each task. As expected, three distinct groups could be formed based on performance on EXPERTise. Cluster 1 comprised those individuals who tended to perform poorly across the EXPERTise tasks, and as such, was designated the low-performance or ‘novice’ cluster. Although, as a group, members of this cluster were faster on the Paired Association tasks than Cluster 2, the lack of variance in their responses suggested that this was probably due to guessing.

Table 3 About Here

Cluster 2 comprised those individuals who demonstrated an intermediate level of performance, generally performing better across the tasks than the members of Cluster 1, but worse than Cluster 3 on all of the tasks. Consequently, the participants in this cluster were described as intermediate-performers, or ‘competent’.

Cluster 3 comprised those individuals who performed at the highest level across the EXPERTise tasks. Since the members of this cluster were generally faster, more accurate, more discriminating and less sequential in the acquisition of information than other participants, they were described as high-performers or ‘experts.’

Comparison of Cluster Experience

One of the key assertions of the present study was that, although novices can be identified by their lack of experience in the domain, experienced-practitioners decompose

into two distinct level of performance that is not necessarily tied to years of experience. Therefore, it was expected that the low-performance group would be relatively less experienced than the mid- and high-performance clusters. The mid- and high-performance groups were not necessarily expected to differ on this dimension.

Table 4 presents the mean years of domain experience and standard deviation for each cluster, as well as the effect size and significance level of the Scheffe post-hoc comparisons between each of the clusters for years of domain experience. Consistent with expectations, the novice cluster was significantly less experienced than the competent, $\eta^2 = .21, p > .05$ and expert clusters, $\eta^2 = .28, p > .05$. In terms of years of experience, the difference between the mid- and high-performance clusters was non-significant, $\eta^2 = .01, p > .05$. This was consistent with the view that that the clusters reflected experienced competence and experienced expertise respectively.

 Table 4 About Here

Comparison between clusters for scenario response

If the three clusters formed reflected three distinct levels of expertise, a significant relationship between cluster membership and diagnostic outcomes should have been observed. Figure 2 illustrates the relationship between cluster membership and the choice of response to the full feature discrimination task scenario.

 Figure 2 About Here

A Chi-square test for independence indicated that there was a significant association between cluster membership and the choice of diagnostic response to the feature

discrimination scenario, $\chi^2(6, 64) = 18.981, p = .004, \phi = .545$. Experts tended to choose the optimal diagnostic response, 'Load shed at 132', as identified by the subject matter expert.

Discussion

The primary aim of the present study was to determine whether performance across four approaches to the assessment of cue utilization performance would decompose into three levels of performance consistent with the novice, competent and expert stages of cognitive skill acquisition described by Ericsson and Lehmann (1996). The assessment tasks were selected on the assumption that diagnostic expertise involves the successful extraction and utilization of cues and that this cognitive skill explains, in part, expert performance. Consequently, expertise was expected to result in consistently superior performance across the diagnostic reasoning tasks.

In general, the results provide substantial support for the assertions underlying the present study. Most importantly, performance clustered into three levels of performance across the four tasks, consistent with the novice, competent, and expert phases described by Ericsson and Lehmann (1996) and Anderson (1983). The expert cluster comprised those participants who were consistently faster, more accurate, showed greater discrimination, and who were less likely to access information in the sequence in which it was presented. By comparison, the competent cluster included those individuals whose performance was slightly lower, but nevertheless, consistent across the tasks. In turn, the competent cluster was clearly distinguished from the novice cluster, which comprised participants who demonstrated the lowest levels of performance. Overall, the groups differed in their diagnostic outcomes, with the expert cluster selecting the optimal response with greater frequency than the competent and novice clusters, suggesting that the groups do in fact genuinely reflect distinct levels of diagnostic performance.

Although a moderate relationship between years of domain experience and task performance was evident, comparisons between the clusters revealed that this relationship was not linear. While the transition from novice to intermediate-performance appears to require a significant increase in years of domain experience, the difference between the intermediate- and expert performance was only trivial. This is broadly consistent with previous research demonstrating that the number of years of experience within a domain is only weakly associated with the progression to expertise (Gray, 2004).

The present results also suggest that expertise can be distinguished from competent and novice performance based on a series of tasks that are dependent on access to, and the utilization of, feature-event relationships or cues (Wiggins, 2006). This is consistent with contemporary perspectives concerning the role of cues in expert diagnostic reasoning and intuition (Kahneman & Klein, 2009). For example, using eye-tracking data, North, Williams, Hodges, Ward, & Ericsson (2009) demonstrated that, in comparison to unskilled soccer players, skilled players tended to extract key features and draw associations between those features to anticipate the destination of the ball following an attacking pass or shot at goal. This feature identification process is widely regarded as an efficient process that reduces cognitive demands, and releases cognitive resources for other tasks (Cellier, et al., 1997; Lansdale, Underwood, & Davies, 2010).

Assessments of Expertise

Many previous assessments of expertise have adopted the expert-novice paradigm in which the performance of experts, classified a priori, is compared to less experienced (typically novice) counterparts (Coderre, et al., 2003; Schraagen, 1993). Although these comparisons are useful, they are predicated on an assumption of expertise, whereby expertise is equated to experience accumulated within the domain. Further, the assessment of performance tends to be based on a relatively limited series of dimensions.

The present study utilized an alternative approach in which: (a) the cohort was carefully selected based on the demand for a particular type of cognitive skill; (b) four assessment tasks were developed on the basis that they assessed different aspects of a fundamental cognitive construct necessary for effective and efficient diagnostic reasoning; (c) a theoretical model of skill acquisition was adopted that justified the delineation of groups; and (d) the groups were delineated on the basis of their performance across the four tasks.

As with all assessments of expertise, the difficulty lies in establishing the validity of expertise in the absence of an unequivocal target or benchmark. One solution is to adopt the approach employed in psychometric testing in which the performance of each individual is assessed against a cohort. By using z-scores, it is possible to make valid comparisons across the cohort and ensure equivalence across the different assessment tasks.

In the case of the present cohort, three distinct clusters emerged that represented three levels of performance, consistent with a three-stage model of skill acquisition (Anderson, 1983). These differences in performance were evident across all four of the assessment tasks, each of which was designed to assess an aspect of cue utilization during diagnostic reasoning (Ratcliff & McKoon, 1995; Weiss & Shanteau, 2003; Wiggins & O'Hare, 1995; Wiggins, et al., 2002). Further, the groups differed in their responses to a more complex diagnostic reasoning task. In combination, this evidence lends weight to the proposition that the groups identified represent distinct populations that differ in their capacity for diagnostic reasoning.

Implications

The results of the present study confirm that three distinct levels of performance can be identified based on performance across the Feature Identification, Paired Association, Feature-Discrimination, and Transition tasks. These three levels of performance are broadly consistent with a three-stage model of cognitive skill acquisition incorporating novice, competent and expert practitioners (Anderson, 1983; Ericsson & Lehmann, 1996).

Importantly, the results indicate that the performance of the competent/intermediate cohort can be differentiated from both the novice and expert cohorts across the range of cue-based tasks.

In addition to confirmation of three levels of performance, it is important to note that all of the assessment tasks in the present study were developed on the assumption that expertise in diagnostic reasoning involves the extraction and utilization of feature-event relationships or cues. Presumably, the utilization of cues enables cognitive processing efficiencies that allow experts to respond more rapidly, with greater discrimination, and with greater depth of processing than their non-expert counterparts (Lipshitz, et al., 2001).

At an applied level, the assessment of expertise based on the extraction and utilization of cues has important implications for the evaluation of training outcomes. For example, it provides a mechanism against which to assess the progression towards expertise. By developing standardized norms for a given domain, it should be possible to establish whether an individual learner is developing diagnostic skills consistent with expectations and/or whether a particular level of performance has been achieved following exposure to training. Measuring independent facets of expert performance, EXPERTise can also be used to identify those aspects of cue utilization and extraction that competent learners are struggling to complete, in order to guide remedial training efforts. Cue-based approaches to training have already met with some success in the aviation context (Wiggins & O'Hare, 2003). The application of this strategy will improve the efficiency and the effectiveness of remedial training and, as a consequence, minimize the costs associated with training interventions.

Finally, at an empirical level, the results support the proposition that domain experts are best identified on the basis of performance (Ericsson & Ward, 2007). The identification of experts on the basis of their within-task performance, rather than their experience per se, should assist with knowledge extraction and empirical comparisons between different levels

of operational performance. The utilization of genuine experts ought to improve the validity of research outcomes involving the expert-novice paradigm and perhaps provide the basis for better understanding the process of cognitive skill acquisition.

Conclusion

The primary aim of the present study was to determine whether the application of four, independent, cue-based assessments of performance collectively differentiated expert, competent, and novice performance in power system control. Overall, performance in the assessment tasks clustered into three groups, thereby providing support for a three-stage model of cognitive skill development based on the acquisition of cue-associations in memory.

The present study was conducted in a context of remote diagnosis using standardized displays. Therefore, it can be presumed that the present results extend to similar domains like air traffic control and rail control. However, EXPERTise may also have utility in other domains, as it can be modified to fit any cognitive skill in which practitioners must acquire visual or auditory information from the environment to make sense of the situation and form a response. Of course, other skills may be of greater significance to expertise in these contexts, like the ability to manage distractions or interruptions, to maintain motivation, or to manage uncertainty. Consequently, further research is necessary to establish the relationship between cue-based reasoning performance and diagnostic accuracy in domains such as medicine, in which high performance requires a combination of cognitive, psychomotor, and psychosocial skills.

The replication of the results of the present study in a domain such as medicine, where studies of expert diagnosis have been hampered by the difficulties associated with reliably identifying expert medical diagnosticians (Norman, Coblenz, Brooks, & Babcock, 1992), would further validate the utility of the EXPERTise tasks. This would also strengthen the

three-stage, cue-acquisition, model of expert performance and ultimately, provide a new solution to an old problem, namely, how to define and identify domain experts.

Acknowledgements

Funding for this project was provided by the Australian Research Council awarded as part of a Linkage Grant with TransGrid (Grant Number: LP0884006).

Key points

- The transition from competence to expertise occurs when associations in memory develop between features of the environment and subsequent events.
- The present study distinguished between the levels of diagnostic expertise using four tasks in which the extraction and utilization of cues may have been advantageous.
- Across the four expert assessment tasks, performance clustered into three-levels, reflecting a three-stage model of cognitive skill acquisition.
- These cue based assessment tasks, collected within EXPERTise, can be used to assess operational performance and to guide remedial training efforts.

Précis

The described study investigated the utility of distinguishing novice, competent and expert performance using four tasks in which the extraction and utilization of cues may be advantageous. The results are discussed in terms of the implications for expert-skill acquisition training and research.

Author bios

Thomas Loveday is a Ph.D. candidate in psychology at Macquarie University, Australia. He received his Masters of Psychology (Organisational) from the University of New South Wales, Australia in 2009.

Mark W. Wiggins is an associate professor of psychology at Macquarie University, Australia. He received his Ph.D. in psychology from the University of Otago, New Zealand, in 2001.

Jemma Harris is a research associate at Macquarie University, Australia. She received her Ph.D. in psychology from the University of Western Sydney in 2010.

Neil Smith is the General Manager of Systems Operations at TransGrid, Australia.

David O'Hare is an associate professor of psychology at the University of Otago, New Zealand. He received his Ph.D. in experimental psychology from the University of Exeter, England, in 1978.

References

- Ackerman, B. P., & Rathburn, J. (1984). The effect of recognition experience on cued recall in children and adults. *Child Development, 55*(5), 1855 - 1864.
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261-295.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2 ed.). Berkeley: University of California Press.
- Cellier, J. M., Eyrolle, H., & Marine, C. (1997). Expertise in dynamic environments. *Ergonomics, 40*, 28 - 50.
- Charness, N. (1976). Memory for chess positions: resistance to interference. *Journal of experimental psychology, 2*, 641 - 653.
- Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education, 37*, 695 - 703.
- de Groot, A. D. (1965). *Thought and choice in chess*. The Hague: Mouton.
- Doane, S. M., Pellegrino, J. W., & Klatzky, R. L. (1990). Expertise in a computer operating system: Conceptualization and performance. *Human-Computer Interaction, 5*, 267 - 304.
- Elo, A. E. (1986). *The rating of chessplayer, past and present* (2nd ed.). New York: Arco Chess.
- Ericsson, K. A. (2004). Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. *Academic Medicine, 79*(10), S70 - S81.
- Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*, 725 - 747.

- Ericsson, K. A., Charness, N., Feltovich, P., & Hoffman, R. (2006). *The Cambridge handbook of expertise and expert performance*. New York: Cambridge University Press.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211 - 245.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaption to task constraints. *Annual review of psychology*, *47*, 273 - 305.
- Ericsson, K. A., & Ward, P. (2007). Occurring Superior Performance of Experts in the Laboratory: Toward a Science of Expert and Exceptional Performance. *Current Directions in Psychological Science*, *16*(6), 346 - 350.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological bulletin*, *51*(4), 327 - 358.
- Flaspöler, E., Hauke, A., Pappachan, P., Reinert, D., Bleyer, T., Henke, N., . . . Beeck, R. O. D. (2009). *The human-machine interface as an emerging risk*. Luxembourg: Office for Official Publications of the European Communities.
- Gale, A. G., & Walker, G. E. (1993). Visual search in breast cancer screening. In D. Brogan, A. Gale & K. Carr (Eds.), *Visual Search 2* (pp. 231 - 228). London: Taylor and Francis.
- Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise, differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, *10*, 42 - 54.
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and instruction*, *20*, 146 - 154.

- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*(6), 515 - 526.
- Klein, G. (1989). Recognition-primed decisions (RPD). *Advances in Man-Machine Systems, 5*, 47 - 92.
- Kolodner, J. (1997). Educational implications of analogy: A view from case-based reasoning. *American Psychologist, 52*, 57–66.
- Lansdale, M., Underwood, G., & Davies, C. (2010). Something overlooked? How experts in change detection use visual saliency. *Applied cognitive psychology, 24*(213 - 225).
- Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of behavioural decision making, 14*, 331 - 352.
- Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological review, 101*, 343 - 352.
- Müller, S., Abernethy, B., & Farrow, D. (2006). How do world-class cricket batsman anticipate a bowler's intention? *The quarterly journal of experimental psychology, 59*, 2162 – 2186.
- Norman, G. R., Coblenz, G. L., Brooks, L. R., & Babcock, C. J. (1992). Expertise in visual diagnosis: A review of the literature. *Academic Medicine, 67*, 78 - 83.
- North, J. S., Williams, A. M., Hodges, N. J., Ward, P., & Ericsson, K. A. (2009). Perceiving patterns in dynamic action sequences: Investigating the processes underpinning stimulus recognition and anticipation skill. *Applied Cognitive Psychology, 23*, 878 - 894.
- O'Hare, D., Mullen, N., Wiggins, M., & Molesworth, B. (2008). Finding the Right Case: The Role of Predictive Features in Memory for Aviation Accidents. *Applied cognitive psychology, 22*, 1163 - 1180.

- Rasmussen, J. (1983). Skills, rules, and knowledge: signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*, 257-266.
- Ratcliff, R., & McKoon, G. (1995). Sequential effects in lexical decision: Tests of compound cue retrieval theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1380 - 1388.
- Reif, F., & Allen, S. (1992). Cognition for interpreting scientific concepts: a study of acceleration. *Cognition & Instruction, 9*, 1 - 44.
- Santos, P. J., & Badre, A. N. (1994). *Automatic chunk detection in human-computer interaction*. Paper presented at the Proceedings of the workshop on advanced visual interfaces, Italy.
- Savelsbergh, G. J. P., Van der Kamp, J., Williams, A. M., Ward, P., & Brogan, D. (2005). Anticipation and visual search behaviour in expert soccer goalkeepers. *Ergonomics, 48*(11-14), 1686 - 1697.
- Schimdt, H. G., & Boshuizen, H. P. A. (1993). Acquiring expertise in medicine. *Educational psychology review, 3*, 205 - 221.
- Schraagen, J. M. (1993). How experts solve a novel problem in experimental design. *Cognitive science, 17*(2), 285 - 309.
- Schrivers, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise Differences in Attentional Strategies Related to Pilot Decision Making. [Article]. *Human Factors, 50*(6), 864-878. doi: 10.1518/001872008x374974
- Schvaneveldt, R. W., Beringer, D. B., & Lamonica, J. A. (2001). Priority and Organization of Information Accessed by Pilots in Various Phases of Flight. *The International Journal of Aviation Psychology, 11*(3), 253 - 280.

- Schyns, P. G. (1998). Diagnostic recognition: task constraints, object information, and their interactions. *Cognition*, *67*, 147 - 179.
- Shanteau, J., & Stewart, T. R. (1992). Why study expert decision making? Some historical perspectives and comments. *Organisational behaviour and human decision processes*, *53*, 95 - 106.
- Strauss, R., & Kirlik, A. (2006). Situation awareness as judgement II: Experimental demonstration. *International Journal of Industrial Ergonomics*, *36*, 475 - 484.
- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning *Cognitive science*, *12*, 257 - 285.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics: 3rd Edition*. New York: HarperCollins College Publishers.
- Vaeyens, R., Lenoir, M., Williams, A. M., & Philippaerts, R. M. (2007). Mechanisms underpinning successful decision making in skilled youth soccer players: an analysis of visual search behaviors. *Journal of Motor Behavior*, *39*(5), 395 - 408.
- Wallis, T. S. A., & Horswill, M. S. (2007). Using fuzzy signal detection theory to determine why experienced and trained drivers respond faster than novices in a hazard perception test. *Accident analysis and prevention*, *39*, 1177 - 1185.
- Watson, R., Stimpson, A., Topping, A., & Porock, D. C. c. a. i. n. A. s. r. o. t. l. J. o. A. N., *39*(5), 421–431. (2002). Clinical competence assessment in nursing: A systematic review of the literature. *Journal of Advanced Nursing*, *39*(5), 421 - 431.
- Weiss, D. J., & Shanteau, J. (2003). Empirical Assessment of Expertise. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*, 104 - 114.
- Wiggins, M. (2006). Cue-based processing and human performance. In W. Karwowski (Ed.), *Encyclopaedia of ergonomics and human factors* (pp. 641 - 645). London: Taylor & Francis.

Wiggins, M., & Bollwerk, S. (2006). Heuristic based information acquisition and decision making among pilots. *Human Factors*, 48, 734 - 746.

Wiggins, M., & O'Hare, D. (1993). A skills bases approach to training aeronautical decision-making. In R. A. Telfer (Ed.), *Aviation Instruction and Training*. Brookfield, VA: Ashgate.

Wiggins, M., & O'Hare, D. (2003). Expert and Novice Pilot Perceptions of Static In-Flight Images of Weather. *The International Journal of Aviation Psychology*, 13(2), 173 - 187.

Wiggins, M., & O'Hare, D. (1995). Expertise in Aeronautical Weather-Related Decision Making: A Cross-Sectional Analysis of General Aviation Pilots. *Journal of Experimental Psychology: Applied*, 1(305 - 320).

Wiggins, M., & O'Hare, D. (2006). Applications of microsimulation in cognitive skills development. In W. Karwowski (Ed.), *Encyclopaedia of ergonomics and human factors* (pp. 3254 - 3259). London: Taylor & Francis.

Wiggins, M., Stevens, C., Howard, A., Henley, I., & O'Hare, D. (2002). Expert, intermediate and novice performance during simulated pre-flight decision-making. *Australian Journal of Psychology*, 54, 162 - 167.