IDENTIFYING DIAGNOSTIC EXPERTS
Measuring the Antecedents to Pattern Recognition

Thomas Loveday[1], Mark Wiggins[1], Marino Festa[2] and David Schell[2]
[1] *Department of Psychology, Macquarie University, Sydney, Australia*
[2] *Children's Hospital at Westmead, Sydney, Australia*

Keywords: Expertise, Diagnosis, Feature Selection, Feature Extraction, Cues, Medicine.

Abstract: Medical expertise is typically denoted on the basis of experience, but this approach appears to lack validity and reliability. The present study investigated an innovative assessment of diagnostic expertise in medicine. This approach was developed from evidence that expert performance develops following the acquisition of cue associations in memory, which facilitates diagnostic pattern-recognition. Four distinct tasks were developed, for which the judicious extraction and selection of environmental cues may be advantageous. Across the tasks, performance clustered into two levels, reflecting competent and expert performance. These clusters were only weakly correlated with traditional methods of identifying domain experts, such as years of experience. The significance of this outcome is discussed in relation to training, evaluation and assessment.

1 INTRODUCTION

1.1 Background

The expertise of medical practitioners has typically been denoted based on cumulative experience in the domain. However, it is apparent that many experienced and qualified practitioners never genuinely attain domain expertise, and instead, only achieve a level of diagnostic performance that could be described as competent.

To explain this observation, Gray (2004) proposed that amongst highly experienced individuals, there may actually be two levels of operators. The levels were presumed to reflect 'competent non-experts', who rely on prior cases and rules (Rasmussen, 1983), and 'genuine experts', who utilise reliable and efficient cognitive shortcuts (Wiggins, 2006). The assertion that experts utilise cognitive shortcuts is consistent with studies that have reported that genuine experts, identified on the basis of diagnostic accuracy rather than experience, are more likely to perform diagnoses using pattern-recognition (Coderre, Mandin, Harasym, & Fick, 2003; Groves, O'Rourke, & Alexander, 2003; Norman, Young, & Brooks, 2007).

In the medical context, pattern recognition is defined as the non-conscious recognition of illnesses based on patterns of symptoms, which primes appropriate responses based on illness scripts in memory (Croskerry, 2009). Under this definition, pattern-recognion, therefore, requires the acquisition of relevant patient features, which are capable of predicting possible outcomes (Lipshitz, Klein, Orasanu, & Salas, 2001; Wiggins, 2006).

The efficiency of expert pattern-recognition suggests that expert practitioners possess highly refined and strong feature-outcome associations in memory (Coderre et al., 2003; Jones, 1992). These 'cue' associations represent an association in memory between the features of the patient and a subsequent outcome or illness (Schimdt & Boshuizen, 1993).

By reducing cognitive load during information acquisition, without sacrificing depth of processing (Sweller, 1988), cue-based pattern recognition allows experts to generate rapid and appropriate responses to environmental stimuli (Wiggins & O'Hare, 2003). For example, in a 'think-aloud' study of gastroenterologists, it was observed that pattern-recognition during diagnosis produced accurate, and seemingly automatic, treatment responses (Coderre et al., 2003).

1.2 The Present Study

Because expert diagnostic performance in medicine

Loveday T., Wiggins M., Festa M. and Schell D. (2012).
IDENTIFYING DIAGNOSTIC EXPERTS - Measuring the Antecedents to Pattern Recognition.
In *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods*, pages 269-274
DOI: 10.5220/0003705902690274

invokes pattern-recognition, it should be possible to distinguish competent individuals from those who have acquired genuine expertise by measuring its component skills. Therefore, the present study proposed distinguishing genuine experts within an experienced population based on their performance on diagnostic tasks in which the selection and extraction of appropriate cues is advantageous.

A battery of cue-based tasks were developed within the software package, EXPERTise (Wiggins, Harris, Loveday, & O'Hare, 2010). EXPERTise was specifically designed to identify expert practitioners by combining four diagnostic tasks:

- *Feature Identification* - a measure of the ability to extract diagnostic cues from the operational environment (Schriver, Morrow, Wickens, & Talleur, 2008);
- *Paired Association* - which assessed the capacity to discern strong feature-event cues from weak feature event cues in the environment (Morrison, Wiggins, Bond, & Tyler, 2009);
- *Feature discrimination* - a measure of the ability to discriminate diagnostic from irrelevant cues in the environment (D. J. Weiss & J. Shanteau, 2003); and the
- *Information Acquisition Task* - assessing the capacity to acquire diagnostic cues from the environment in a strategic, non-linear pattern (Wiggins, Stevens, Howard, Henley, & O'Hare, 2002).

It had already been established that the EXPERTise tasks could consistently and accurately distinguish the performance of novice, competent and expert network diagnosticians in the context of power control (Loveday, Wiggins, Harris, Smith, & O'Hare, submitted). The present study had the distinct aim of determining the utility of EXPERTise in distinguishing competent non-experts from genuine experts within an experienced sample of medical practitioners.

Because each of the four tasks used in the present study was selected to assess independent facets of the broader construct of pattern-recognition based diagnosis, *it was hypothesised that performance amongst experienced practitioners would cluster into two levels across the tasks*, consistent with the predictions of Gray (2004). Because experience is only weakly associated with expert skill acquisition, performance on the tasks assessing expert performance, *were not expected to correlate significantly with measures of domain experience*.

2 METHOD

2.1 Participants

Fifty paediatric intensive care unit staff were recruited. Twenty three were male and twenty seven were female. They ranged in age from 30 to 63 years with a mean of 42.3 years (SD = 8.3). The participants had accumulated between 3 and 26 years of experience within paediatric critical care, with a mean of 9.8 years (SD = 6.9).

2.2 Measures

2.2.1 Demographic Survey

In addition to basic demographics, general and specific experience in the domain were recorded.

2.2.2 EXPERTise

EXPERTise (Wiggins, Harris, Loveday, & O'Hare, 2010) is a 'shell' software package designed to record performance across four cue-based expert reasoning tasks. EXPERTise was specifically designed so that these tasks could be customized to match stimuli used in the domain.

2.3 Stimuli

Cognitive interviews were conducted with two paediatric intensive care practitioners to develop the stimuli used in the present study. These practitioners were selected on the basis of peer recommendation. The information derived from the subject-matter experts was restructured into several scenarios that identified feature and outcome pairs that were available for patient diagnosis. These pairs and scenarios were validated in an untimed pilot test. The scenarios formed the basis of the stimuli used within the EXPERTise tasks. See Figure 1 For an example of the stimuli.

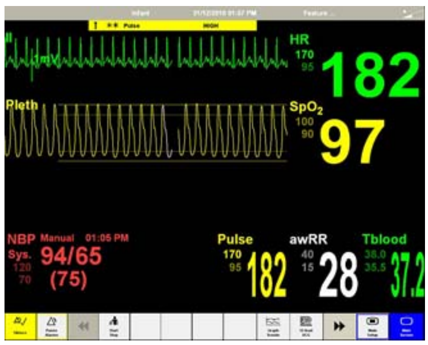

Figure 1: Example patient bedside monitor output.

2.3.1 Feature Identification Task

The feature identification task had two forms. In the first, the participants were presented with a patient bedside monitor displaying an abnormal parameter that indicated that the patient was in a critical condition. The participants were asked to click on the abnormal parameter. In the second form, the bedside monitor was 'flashed' for 1.5 seconds, and the participant was asked to identify the abnormal parameter from one of four options. For both forms, response times were recorded and aggregated across items to yield a mean response time. Accuracy was also recorded and totalled into a single accuracy score.

2.3.2 Paired Association Task

The paired association task also had two forms. In both, two domain-relevant phrases were flashed on-screen, either sequentially (Form 1) or simultaneously (Form 2) for 1.5 seconds. The participant was asked to rate the relatedness of the two phrases on a six-point scale.

Response latencies were recorded and aggregated across items to yield a mean reaction time for each participant. The association ratings were also aggregated into a single 'discrimination' metric, based on the mean variance of the participants' responses.

2.3.3 Feature Discrimination Task

The feature discrimination task measured expert discrimination between sources of information during decision-making. The task presented the participant with a patient bedside monitor output and a short written scenario description. On a subsequent screen, the participants were asked to choose an appropriate response to the scenario from eight treatment options. The participants then rated, on a six-point scale, the utility of nine individual types of information in informing their decision. These ratings were aggregated into a single discrimination metric based on the variance of the participant's ratings.

2.3.4 Information Acquisition Task

The information acquisition task consisted of a single scenario accompanied by a patient bedside monitor output. The scenario was intentionally vague and thus, forced to participant acquire additional information as provided in a list of information screens. The participants then selected an appropriate diagnosis and response from four treatment options. The order in which the information screens were accessed was recorded. This was converted to a single metric based on the ratio of screens accessed in sequence over the total number of screens accessed.

2.4 Procedure

Conducted in groups of five, participants were briefed on the purpose of the study and then asked to sign a consent form if they wished to continue. They then completed the demographics questionnaire and EXPERTise via laptops.

3 RESULTS

3.1 Correlations with Experience

To investigate the relationship between measures of experience and each task within EXPERTise, bivariate correlations were undertaken between years of experience (both domain general and domain specific) and performance on the EXPERTise tasks. Consistent with expectations, measures of experience, both general and specific, yielded only weak to moderate Pearson correlations with performance on the EXPERTise tasks, $r \leq 0.33$, $p < 0.05$.

3.2 Cluster Models

The primary aim of the present investigation was to determine the feasibility of identifying expert practitioners using tasks in which pattern recognition was advantageous. Because the sample comprised qualified individuals, it was expected that performance would cluster into two groups, reflecting competence and expertise.

Table 1: Participant cluster means.

Measure	Competent Mean (SD) n = 24	Expert Mean (SD) n = 26	Overall Mean (SD) N = 50
FID RT	11.1 (4.4)	7.7 (3.0)	9.2 (4.1)
FID Acc	5.3 (2.1)	6.7 (1.7)	6.0 (2.0)
PAT1 RT	6.0 (2.2)	4.6 (1.3)	5.3 (1.9)
PAT1 Var	1.5 (0.6)	2.4 (0.8)	2.0 (0.8)
PAT 2 RT	4.3 (1.7)	3.6 (1.1)	3.9 (1.4)
PAT 2 Var	1.2 (0.7)	1.8 (0.7)	1.5 (0.7)
FDT Var	2.72 (2.8)	4.5 (3.2)	3.7 (3.1)
IAT Ratio	0.91 (0.17)	0.63 (0.42)	0.8 (0.4)

SD = Standard Deviation; FID = Feature Identification; PAT = Paired Association Task; FDT = Feature Discrimination; IAT = Information Acquisition Task. RT = Reaction Time; Acc = Accuracy; Var = Variance.

Table 1 presents the results of a K-Means cluster analysis. As expected, two distinct groups formed based on performance across the EXPERTise tasks.

Cluster 1 (n = 24) comprised those individuals who, whilst qualified, demonstrated a lower level of performance across the EXPERTise tasks in comparison to the members of Cluster 2. Therefore, the participants in this cluster were described as ‘experienced non-experts’.

Cluster 2 (n = 26) 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 their acquisition of information, they were described as ‘genuine experts.’

4 DISCUSSION

The aim of the present study was to determine whether four measurements of pattern recognition could, when combined, distinguish competent from expert paediatric healthcare practitioners within an experienced sample. Because the judicious selection and extraction of cues was advantageous in each of the tasks, it was expected that paediatric experts would demonstrate consistently superior performance.

The results of the present study are consistent with expectations that the EXPERTise tasks could consistently distinguish between competent and expert practitioners within an experienced sample. Performance across the four assessment tasks clustered into two levels, with the genuine expert cluster significantly outperforming the competent cluster on each task.

As expected, performance in the tasks was not strongly correlated with domain experience. This outcome is consistent with prior research (Coderre et al., 2003; Groves et al., 2003; Norman et al., 2007), and thus, further highlights the limitations of this approach as a means of identifying expert diagnosticians in paediatric healthcare. There is an increasingly strong case to be made that experience is only weakly associated with the progression to diagnostic expertise (Gray, 2004), indicating that other indicators may be preferable.

4.1 Implications for Theory and Research

Many prior studies of expert diagnosis have attempted to identify medical experts in advance, usually on the basis of experience (Blignaut, 1979; Coderre et al., 2003; O’Hare, Mullen, Wiggins, & Molesworth, 2008; Simon & Chase, 1973; Wallis & Horswill, 2007). Although these comparisons can be useful, they are based on the assumption that there is a linear relationship between experience and diagnostic performance. However, in the present study, performance in four tasks, all of which have been linked to expertise, was only weakly associated with experience. Therefore, while experience may be a necessary precursor to expert diagnostic performance, it is not sufficient.

The present results suggest that when investigating diagnostic performance in medicine, expertise should not be operationalised as experience in the domain. The present study supports an alternative approach, based on the assessment of pattern recognition during domain-relevant tasks. This solution, whereby the performance of each individual is assessed against a cohort, makes it possible to make valid comparisons between individuals.

In the present cohort, comprising experienced practitioners, two distinct clusters emerged that appear to represent two distinct levels of performance. These levels were consistent with the distinction made by Gray (2004) between competent and expert practitioners. Moreover, these differences in performance were consistent across all four assessment tasks, each of which was designed to assess an independent dimension of expert pattern recognition (Morrison et al., 2009; Ratcliff & McKoon, 1995; D. J. Weiss & J. Shanteau, 2003; Wiggins & O’Hare, 1995; Wiggins et al., 2002).

The identification of medical experts on the basis of their performance, rather than their experience, should assist with studies of feature extraction, pattern recognition and empirical comparisons between different levels of diagnostic performance. Further, the identification of genuine experts ought to improve the validity of research outcomes involving the observation of expert performance, and perhaps, provide the basis for a better understanding of the process of cognitive skill acquisition.

4.2 Implications for the Field

At an applied level, the assessment of expertise based on feature extraction, cue utilisation, and pattern recognition has important implications for evaluation. In particular, it provides a method for assessing the progression towards medical expertise.

With the development of standardized norms, it should be possible to determine 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 specialist training. By assessing four components of expert pattern recognition, EXPERTise can also be used to identify those component skills of pattern recognition that experienced competent practitioners are struggling to acquire. This information can then guide remedial training efforts. Such cue-based approaches to training have already met with some success in other domains, including aviation (Wiggins & O'Hare, 2003) and mining (Blignaut, 1979).

The nature of the assessment tasks' is such that they assess independent skills, each of which contribute to expert pattern recognition and diagnosis. Therefore, if performance is weaker on one or more of the tasks, it will be possible to identify the specific area of deficiency and thereby better target interventions. The application of this strategy can be used to improve the efficiency and the effectiveness of remedial medical training and, as a consequence, minimize the costs associated with training interventions.

5 CONCLUSIONS

The present study was designed to determine whether four independent assessments of expert pattern recognition could, collectively, distinguish competent from expert practitioners within a qualified sample of healthcare practitioners. Overall, performance on all four assessment tasks successfully differentiated the two groups, whereby qualified staff could be divided into competent and expert practitioners based on their capacity for pattern recogniton, and cue extraction and utilisation.

The successful replication of the results of Loveday, et al. (submitted) in a dissimilar domain demonstrates the utility of the EXPERTise tasks, and the importance of pattern recognition in expert performance generally. In time, it may also provide a method for determining whether experienced practitioners are developing expertise at a rate that is consistent with their peers. Individuals' who perform at an unsatisfactory level may benefit from remedial medical training. It is expected that this combination of progressive assessment and remedial training may reduce the rate of error in medicine through the increased diagnostic expertise of practitioners.

ACKNOWLEDGEMENTS

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).

REFERENCES

Blignaut, C. J. (1979). The perception of hazard: I. Hazard analysis and the contribution of visual search to hazard perception. *Ergonomics, 22*, 991 - 999.

Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education, 37*, 695 - 703.

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine, 84*(8), 1022 - 1028.

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.

Groves, M., O'Rourke, P., & Alexander, H. (2003). The clinical reasoning characteristics of diagnostic experts. *Medical Teacher, 25*, 308 - 313.

Jones, M. A. (1992). Clinical reasoning in manual therapy. *Physical therapy, 72*, 875 - 884.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. Journal of behavioural decision making. *Journal of behavioural decision making, 14*, 331 - 352.

Loveday, T., Wiggins, M., Harris, J., Smith, N., & O'Hare, D. (submitted). An Objective Approach to Identifying Diagnostic Expertise Amongst Power System Controllers.

Morrison, B. W., Wiggins, M. W., Bond, N. W., & Tyler, M. D. (2009). *Examining cue recognition across expertise using a computer-based task.* Paper presented at the NDM9, the 9th International Conference on Naturalistic Decision Making, London.

Norman, G. R., Young, M., & Brooks, L. (2007). Non-analytical models of clinical reasoning: the role of experience. *Medical education, 41*(12), 1140 - 1145.

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.

Schimdt, H. G., & Boshuizen, H. P. A. (1993). Acquiring expertise in medicine. *Educational psychology review, 3*, 205 - 221.

Schriver, 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.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American scientist, 61*, 394 - 403.

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning *Cognitive science, 12*, 257 - 285.

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.

Weiss, D. J., & Shanteau, J. (2003). Empirical Assessment of Expertise. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 45*, 104 - 114.

Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. *Human Factors, 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., & 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. (2003). Weatherwise: an evaluation of a cue-based training approach for the recognition of deteriorating weather conditions during flight. *Human Factors, 45*, 337 - 345.

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.