This is the author version of an article published as:


**Access to the published version:** [http://dx.doi.org/10.1177/0018720812448475](http://dx.doi.org/10.1177/0018720812448475)

**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

The capability of static and dynamic features to distinguish competent from genuinely expert practitioners in pediatric diagnosis

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

Festa, M., Schell, D.
Children’s Hospital at Westmead, Sydney, Australia

Author Note

Thomas Loveday, Department of Psychology, Macquarie University; Mark Wiggins, Department of Psychology, Macquarie University; Ben Searle, Department of Psychology, Macquarie University; Marino Festa, Children’s Hospital at Westmead; David Schell, Children’s Hospital at Westmead.

This research was supported in part by grants from the ‘Australian Research Council’ 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, competence, novice, assessment

Word count: 5086
Précis

The present study distinguished competent non-expert and expert practitioners within an experienced and qualified sample of pediatric intensive care staff, based on performance in a series of tasks in which the extraction of static or dynamic cues was advantageous.
Abstract

Objective: The present study describes the development of a new, more objective, method of distinguishing experienced competent non-expert from expert practitioners within pediatric intensive care.

Background: Expert performance involves the acquisition and utilization of refined feature-event associations (cues) in the operational environment. Competent non-experts, although experienced, possess rudimentary cue-associations in memory. Thus, they cannot respond as efficiently or reliably as their expert counterparts, particularly when key diagnostic information is unavailable, such as that provided by dynamic cues.

Method: This study involved the application of four distinct tasks in which the use of relevant cues could be expected to increase both the accuracy and efficiency of diagnostic performance. These tasks included both static and dynamic stimuli that were varied systematically. Fifty experienced pediatric intensive staff took part in the study.

Results: The sample clustered into two levels across the tasks with participants who performed at a consistently high level throughout the four tasks labeled experts and participants who performed at a lesser level throughout the tasks labeled competent non-experts. The groups differed in their responses to the diagnostic scenarios presented in two of the tasks and their ability to maintain performance in the absence of dynamic features.

Conclusion: Experienced pediatricians can be decomposed into two groups on the basis of their capacity to acquire and utilize cues; these groups differ in their diagnostic accuracy and in their ability to maintain performance in the absence of dynamic features.

Application: The tasks may be used to identify practitioners who are failing to acquire expertise at a rate consistent with their experience, position or training. This information may be used to guide targeted training efforts.
The capability of static and dynamic cues to distinguish competent from genuinely expert practitioners

Concerted efforts have been undertaken to improve the diagnostic performance of intensive care unit staff, with the goal of reducing the rate of patient misdiagnosis (Litchfield, Ball, Donovan, Manning, & Crawford, 2010). Stemming from the observation that accurate and efficient diagnostic performance only becomes stable with a high-degree of specialist expertise within medicine (Patel & Groen, 1991), extensive training and clinical experience is provided to practitioners to facilitate the acquisition of expert knowledge and skills (Ericsson & Lehmann, 1996). However, the efficacy of such programs is debatable, with repeated observations that the diagnostic performance of experienced and qualified medical practitioners is little better than novice graduates (Norman, Coblentz, Brooks, & Babcock, 1992).

Part of the problem relates to the conceptualization of expertise within healthcare, whereby individuals with relevant experience and qualifications are designated as experts in their domain (Coderre, Mandin, Harasym, & Fick, 2003). 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 (Gray, 2004; Groves, O’Rourke, & Alexander, 2003; Loveday, Wiggins, Harris, Smith, & O’Hare, submitted; Schneider, 1985).

There is, therefore, utility in being able to distinguish between competent non-experts and genuine experts for the purposes of remedial training and program evaluation. Thus, the general aim of the present study was the development of a tool capable of distinguishing competent from expert practitioners within an experienced and qualified population of intensive care unit staff.

An objective conceptualization of expertise in healthcare
The present study bases its conceptualization of expertise in medicine on the distinction made by Croskerry (2009) between System 1 and System 2 decision-making. System 1 is an approach to diagnosis based on clinical intuition, which facilitates rapid and accurate diagnosis by non-consciously matching patterns of symptoms to appropriate treatment heuristics (Croskerry, 2009; Schmidt, Norman, & Boshuizen, 1990). By comparison, System 2 is a slow and deliberate process, whereby practitioners form an initial diagnosis when presented with a range of symptoms, and then test it by deducing what further symptoms should be present (Croskerry, 2009; Gilhooly, 1990). It is well established that expert medical practitioners are much more likely than their non-expert peers to engage in System 1 type pattern-recognition (Coderre et al., 2003; Groves et al., 2003; Norman, 2005; Young, Smith, Guerlain, & Nolley, 2007), although non-experts may also do so when presented with typical or common cases (Croskerry, 2009).

Klein’s (1989) recognition-primed model of decision-making (RPD) proposed that experts engage in pattern recognition through the non-conscious extraction and utilization of combinations of cues and their associations in memory. In the healthcare context, these cues would represent associations in memory between diagnostic features and patient events (Schmidt & Boshuizen, 1993). Cues facilitate efficient and accurate diagnosis by priming the most appropriate response and may also provide a basis to discriminate relevant from less relevant information in the environment by focusing attention towards a particular task (Rasmussen, 1983).

There is extensive evidence to indicate that, although all levels of diagnostic performance require the acquisition of task-relevant cues, expert performance is characterized by the capacity to draw on stronger and more nuanced associations in memory – producing consistent, accurate and efficient diagnoses (Cellier, Eyrolle, & Marine, 1997; Lipshitz, Klein, Orasanu, & Salas, 2001; Müller, Abernethy, & Farrow, 2006; Schriver, Morrow,
Wickens, & Talleur, 2008; Wiggins, 2006). By contrast, competent practitioners, with only weak or rudimentary associations, may focus on only the most salient or common cues during diagnosis, rather than the most relevant given the circumstances, resulting in less efficient and less accurate diagnostic performance (Cellier et al., 1997; Lipshitz et al., 2001; Schriver et al., 2008; Wiggins, 2006). Consequently, the present study distinguishes the stages of expert skill acquisition in terms of the capacity to extract and utilize diagnostic cues (Cellier et al., 1997; Lipshitz et al., 2001; Schriver et al., 2008).

On the basis of this conceptualization of expertise, it should be possible to distinguish competent non-experts from genuine experts within intensive care, based on diagnostic tasks in which the extraction and utilization of cues is advantageous. The Expert Intensive Skills Evaluation (EXPERTise) is an assessment tool that has been designed to measure discrete components of cue extraction and/or utilization using a series of four tasks, namely the Feature Identification Task, the Paired Association Task, the Feature Discrimination Task, and the Transition Task (Wiggins, Harris, Loveday, & O’Hare, 2010). The utility of EXPERTise has been established in a number of domains, including power system control (Loveday et al., submitted).

The Feature Identification task (FID) measures the ability to extract diagnostic features from the operational environment and recall those features when required (Loveday et al., submitted). It was developed from observations that experts are typically primed by a small subset of cues in the environment that are associated with critical events (Müller et al., 2006; Schriver et al., 2008). These cues both generate the initial recognition of a critical condition, and prime/restrict the search for further cues (Rasmussen, 1983). Consequently, these cues produce greater efficiency in diagnosis, while maintaining a high level of diagnostic accuracy (Coderre et al., 2003; Croskerry, 2009; Schyns, 1998). Modified for the medical domain, the feature identification task can be used to measure the speed and
accuracy with which a practitioner is able to identify patient features indicative of critical illness. Expert practitioners would be expected to respond faster and more accurately than their non-expert peers.

The *Paired Association* task (PAT) measures the perceived strength of association between features and outcomes in the domain by presenting domain-relevant word pairs and measuring response latency or ratings of relatedness (Morrison, Wiggins, Bond, & Tyler, 2009). Experts who possess more nuanced and refined feature event cue-associations in memory respond faster and with greater variance to pairs than non-experts, for whom all pairs are similarly associated (Ackerman & Rathburn, 1984; Schvaneveldt, Beringer, & Lamonica, 2001). This task can be easily modified for medicine, by using diagnostic words and symptoms drawn from patient reports and charts.

The *Feature Discrimination* task (FDT) measures the perceived utility of individual features within a diagnostic scenario. It is based on the Cochran-Weiss-Shanteau index, which proposed that expertise can be measured using a ratio of discrimination over consistency, where discrimination is calculated as the variance in ratings for different features within a scenario and consistency as the variance in ratings for the same feature between scenarios (Weiss & Shanteau, 2003). The feature discrimination task, which utilizes more complex, less standardized scenarios than those described by Weiss and Shanteau (Weiss & Shanteau, 2003), uses only the discrimination component of that formula. The task can be modified to closely reflect the types of information available in patient handover notes. Expert medical practitioners would be expected to demonstrate greater discrimination in their ratings of utility during a diagnostic scenario, resulting in greater variance of ratings.

The *Transition* task (TT) measures the sequence with which features are acquired during diagnosis (Wiggins, Stevens, Howard, Henley, & O’Hare, 2002). The task is based on the observation that experts and non-experts differ in the sequence in which they acquire
diagnostic information. Typically, non-experts will acquire information in the sequence that it is presented, whereas experts, who are guided by cue-associations between features and events, acquire information in a sequence based on relevance and context (Wiggins & Bollwerk, 2006; Wiggins & O’Hare, 1995). The transition task operationalizes the sequence of information acquisition by presenting a list of feature categories. When clicked, these categories provide more detailed information. The sequence in which the category is accessed can be operationalized as the ratio of categories inspected in order of appearance over the total number of categories inspected prior to diagnosis. Therefore, expert practitioners would be expected to demonstrate a lower ratio of categories accessed in the sequence they were presented than non-expert practitioners. This task can be modified to closely reflect the types of information available in a patient report albeit with a novel structure. It can also be appended by a patient beside-monitor output.

**The present study**

Although previous applications of EXPERTise have been limited to power control, there are fundamental similarities between patient diagnosis in medicine and electrical network fault diagnosis. Power control operators monitor the power network using a standardized display, which both alerts the operator to line faults, and facilitates fault diagnosis through prominent visual cues (Loveday et al., submitted). While the cues used by intensive care staff are arguably more complex and dynamic, they too must extract relevant diagnostic information from standardized bedside monitors (Norman, et al. 1992). Therefore, it may be possible to identify genuinely expert healthcare staff using EXPERTise, in much the same way that EXPERTise has been used to identify expert network operators.

A limitation of Loveday, et al. (submitted) was that it made use of static stimuli, in which the available cues may have lacked information inherent to dynamism, such as the rate of change, the magnitude of change, and the presence or absence of change (Tversky,
Morrison, & Betrancourt 2002). Likewise, most of the studies of expert healthcare diagnosis have opted to present diagnosticians with static patient summaries, listing patient symptoms, history and demographics (Coderre et al., 2003).

This is problematic for research on expert diagnosis within intensive care, where time-pressures mean that practitioners frequently assess the efficacy of their treatment response based on observable changes in conditions (Rasmussen, 1993). It is only in highly specific circumstances that practitioners attempt to infer patient changes in conditions from static information like x-rays.

A failure to replicate actual field conditions may significantly impair performance (Shanteau, 1992). The extent of this impairment may depend on whether the individual is a competent non-expert or genuine expert. Compared to their expert counterparts, competent non-experts have greater difficulty interpreting dynamic environments using static cues in isolation. For example, Lowe (2001) had participants predict and draw meteorological markings on a map of Australasia based on a static meteorological map of the Australian mainland. It was observed that non-experts produced very simple forecast maps compared to those produced by experts.

Due to the potential significance of dynamic features in diagnostic performance, there is a need to assess performance in the presence or absence of dynamic cues. By using both static cues and dynamic cues, the present study was designed to examine differences in performance between expert and competent non-expert intensive care staff within an experienced sample.

**Aims and hypotheses**

The specific aim of the present study was to establish whether four tasks, designed to assess independent aspects of diagnostic reasoning based on the utilization and extraction of both static and dynamic cues, were able to differentiate experienced competent non-expert
and expert staff in pediatric intensive care. Three of the tasks, feature identification, feature-discrimination, and the transition task, included both static and dynamic stimuli to examine the extent to which diagnosticians based their decisions on information such as the rate of change and the direction of change.

Because experts possess nuanced cue-associations in memory that direct their performance, they were expected to demonstrate consistently superior performance on those reasoning tasks that required the judicious extraction and utilization of cues. Therefore, it was hypothesized that:

1. Performance across the EXPERTise tasks would cluster into two levels within an experienced sample.

2. Reflecting competent non-experts and genuine experts, the clusters would differ in terms of performance in each of the EXPERTise tasks. Experts, compared to non-experts, were expected to demonstrate:
   a. Both shorter response intervals and increased accuracy during the feature identification task;
   b. Shorter response intervals and larger variance of ratings of association between pairs in the paired association task;
   c. Larger variance in ratings of utility of individual features during the feature-discrimination task; and
   d. A smaller proportion of information screens accessed sequentially during the transition task.

3. Overall performance in the tasks would be associated with the choice of response to the EXPERTise tasks (Feature Discrimination and the Transition Tasks) presenting full diagnostic scenarios, with the expert cluster tending to select the optimal response to the scenario.
If competent non-experts were unable to impose dynamism on static stimuli, their performance should be worse on static than dynamic items. This effect should be less pronounced for genuine experts. Thus, it was also expected that:

4. Across the EXPERTise tasks, competent non-experts would perform significantly worse when only static, but not dynamic, stimuli were available. Members of the expert cluster would perform similarly, regardless of stimuli type.

Method

Participants

The participants represented a convenience sample recruited with the assistance of the Children’s Hospital at Westmead (CHW) during the World Federation of Pediatric Intensive and Critical Care Societies (WFPICCS) Meeting, held in Sydney, March 2011. The participants comprised 50 pediatric intensive care unit staff, 34 registered pediatric intensivists and 16 registered pediatric nurses. Twenty-three were male and 27 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 beyond residency within pediatric critical care, with a mean of 9.8 years (SD = 6.9).

Stimuli Development

To develop the stimuli, cognitive interviews were conducted with two subject matter experts from the pediatric intensive care unit within the Children’s Hospital at Westmead. The subject matter experts were selected on the basis of their position, peer reference and the length of their experience practicing in the domain. Initially, the subject matter experts were interviewed as a pair. They were asked to describe a typical, but rapidly deteriorating patient, and to list the features that may have been available, even if they believed that the information was irrelevant to their final decision. This approach was taken to circumvent the
possibility that such a skilled practitioner may have been unaware of the cues that guided decision-making (Kahneman & Klein, 2009).

The features identified were used to develop a scenario template, which included a catalogue of feature categories that an individual may or may not consider during a complex diagnostic task within pediatrics. The subject matter experts were asked to construct scenarios from real incidents, describing the patient in terms of the features listed in the template. It was emphasized that the incidents reported should be complex due to the potential for uncertainty, and where the subject matter experts felt that their expertise and experience made a critical difference to the outcome. Four incidents were organized into timelines reflecting, as accurately as possible, the actual sequence and timing of events. Two of the scenarios were adapted into the Feature Discrimination tasks (See Appendix A: The Feature Discrimination Task for an example) and two into the Transition Tasks (See Appendix B: The Transition Task for an example). The remaining incidents were used to develop a catalogue of features, which were identified by the subject-matter-experts. These features were then paired randomly to form the items in the Paired Association task (See Appendix C: Paired Association Task Items).

The subject matter experts also provided critical patient vital sign ranges, which were entered as variables into Laerdal’s SimBaby™, a pediatric infant patient monitor simulation, to generate static patient monitor screen-captures, as shown in Figure 1, and dynamic patient monitor videos utilizing a flashing signal to draw attention to critical patient vitals. These images and scenarios were validated during an untimed pilot test, whereby five senior intensive care practitioners within the PICU/CICU at the Children’s Hospital at Westmead identified the abnormal parameters for each of the images, and nominated the optimal response options for the scenarios. Only those items with 100% agreement were retained. The resulting scenarios, images and videos were used across the four EXPERTise tasks.
Stimuli

In the present study, the Feature Identification task (FID) incorporated four stages. In the first stage, the static cue condition, the participants were presented with a static screen capture. In the second stage, the dynamic cue condition, the participants were presented with a video recording of a patient’s bedside monitor. For both the static and the dynamic condition, the monitor output displayed a single abnormal parameter, indicative of a patient in a critical condition. The speed with which the participants were able to identify and ‘click’ on the abnormal parameter was recorded. Due to the simplicity of the task, it was expected and observed that all participants would correctly identify the abnormal parameter for all items.

In the third stage, the static cue condition, the participants were shown a static monitor screen capture for 1.5 seconds. In the fourth stage, the dynamic cue condition, the participants were shown a video recording for 1.5 seconds. During both stages, an abnormal parameter occurred. On a subsequent screen, they were asked to identify the abnormal parameter from five possible options and their responses were recorded.

The Paired Association task (PAT) incorporated two stages. In the first stage, participants were presented with pairs of feature and event phrases sequentially. For example, the participants might have been shown the phrases ‘Breathing’ and then ‘Sedation’. The stimuli were presented as text pairs for 1.5 seconds. The second stage presented the feature and event pairs simultaneously on the screen for two seconds. The participant was asked to rate the perceived strength of the relationship between the feature-event pairs, which was recorded, together with response latency.

The Feature Discrimination task (FDT) included two scenarios. Both included a brief scenario description and a patient monitor output that indicated the need for further action.
However, the first presented a static screen capture, whereas the second presented a dynamic video recording. Participants were asked to assess the patient and then, on the same screen, select one of eight possible treatment responses that included a ‘Do nothing’ option. On the following screen, they rated, on six-point scales, the perceived diagnostic utility for each of nine individual features of the scenario. The participants’ response to the scenario, and their ratings of each feature, were recorded.

The Transition task (TT) also included two scenarios. Each scenario provided the gender and age of the patient, to provide context, and a patient monitor output. In the first scenario, the monitor output was provided as a static image. In the second, the output was a dynamic video recording. The monitor output and written description were intentionally vague, and thus, were insufficient to make a diagnosis. This necessitated the participants’ acquisition of additional information, which was presented as a list of information categories. Each information category, when clicked, revealed more detailed information. Each label was a term chosen to ensure that its contents were clear to participants and this was verified during pilot testing. The sequence in which these categories were accessed was recorded.

**Procedure**

Conducted in groups of approximately two to five participants, participants 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 their pediatric experience and perceived workload, and then began the EXPERTise test battery via laptop computer. Noise-cancelling headphones were provided to reduce the impact of ambient noise. On completion of EXPERTise, the participants were debriefed.

**Results**

**Correlations with Experience**
To investigate the relationship between measures of experience and each of the tasks within EXPERTise, Bonferroni adjusted bivariate correlations were undertaken between years of experience in the domain (Years in PICU/CICU), years of senior experience (If Consultant in Charge of PICU), workload (Hours in typical week, Patients admitted to unit per year), and performance on the EXPERTise tasks.

There were no statistically significant correlations between measures of experience and task performance, with the exception of years of PICU experience, which was moderately correlated with sequential PAT reaction times, \( r = .33, p < .05 \). There was also a moderate correlation between workload, as measured by consult hours in a typical week, and ratings variance in the static Feature-Discrimination task, \( r = .40, p < .05 \).

**Cluster analysis**

The primary aim of the present study was to investigate the feasibility of classifying pediatric intensive care unit staff into a limited number of groups on the basis of observed levels of performance across the EXPERTise tasks. Because the sample comprised experienced individuals, it was expected that performance would decompose into two groups, reflecting competence and expertise. Therefore, a K-means cluster analysis was undertaken in SPSS, with \( K = 2 \) clusters.

Table 1 presents the results of the cluster analysis, the mean score for each cluster on each of the outcome variables, and the order (ranking) in which the groups performed on each task. As expected, two distinct groups could be formed based on performance across the EXPERTise tasks.

Cluster 1 (\( n = 24 \)) comprised those individuals who, although experienced, demonstrated a lower level of performance across the EXPERTise tasks than the members of Cluster 2. Therefore, the participants in this cluster were described as ‘competent 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 the acquisition of information than other participants, they were described as ‘experts.’

Comparison between clusters for scenario response

If the clusters formed reflected two distinct levels of expertise, a significant relationship between cluster membership and diagnostic outcomes should have been observed. However, because the clusters were not significantly different in the dynamic forms of the feature discrimination and transition tasks, only the diagnostic outcomes for the static forms of the feature discrimination and transition tasks were considered. Figures 2 and 3 illustrate the relationship between cluster membership and the choice of response to the feature discrimination and transition tasks respectively.

A Chi-square test for independence indicated that there was a significant association between cluster membership and the selection of the optimal diagnostic response to the static transition task scenario, \(\chi^2(1, 50) = 5.77, p = .02, \phi = .34\). Experts tended to choose the optimal diagnostic response, ‘Start a norepinephrine (noradrenaline) infusion’, as identified by the subject matter experts and pilot testing. The association between cluster membership and selection of the optimal diagnostic response in the static feature discrimination scenario was not significant \(\chi^2(1, 50) = 1.94, p = .16, \phi = .20\).

Dynamic vs. static stimuli
A mixed between-within subject’s multivariate analysis of variance was conducted to assess the relationship between stimuli type (static, dynamic) and EXPERTise classification (competent non-expert, expert) on performance in each of the tasks. There was no significant main effect for stimuli type, with $\text{Wilks Lambda} = .99$, $F(2, 48) = 4.465$, $p = .821$, $\eta^2 = .001$, suggesting no overall difference in performance due to the presence or absence of dynamism. However, there was a significant interaction effect between EXPERTise classification and stimuli type, $\text{Wilks Lambda} = .92$, $F(2, 48) = 4.465$, $p = .04$, $\eta^2 = .09$, with members of the competent cluster performing substantially better when dynamic stimuli were available. By contrast, members of the expert cluster performed similarly regardless of stimuli type.

**Discussion**

The primary aim of the present study was to determine whether four diagnostic tasks, in which the extraction and utilization of cues was advantageous, were able to differentiate non-expert and expert practitioners within an experienced sample of pediatric intensive care staff. These tasks had already distinguished novice, competent and expert practitioners in power control (Loveday et al., submitted), and were considered appropriate for the assessment of diagnostic expertise within an experienced pediatric intensive care sample, based on prior evidence that experienced practitioners only transition to expertise with the acquisition of cue-associations in memory (Cellier et al., 1997; Lipshitz et al., 2001; Schriver et al., 2008; Wiggins, 2006). Therefore, expertise was expected to result in consistently superior performance across the tasks.

A secondary aim of the study was to examine the relationship between expertise and performance in the presence and absence of dynamic information. It was expected that, in the absence of dynamic features, competent non-experts would perform at a lower level than would otherwise be expected. Thus, it was hypothesized that the difference between
competent and expert practitioners would be greatest for those tasks that presented static, rather than dynamic, stimuli.

The first hypothesis, that the EXPERTise tasks could distinguish competent and expert practitioners within an experienced sample was supported. Performance clustered into two levels across all four diagnostic assessment tasks. The expert cluster comprised those participants who were faster, more accurate, more discriminating in their selection of diagnostic cues, and who were the less likely to inspect information in the order in which it was presented. By comparison, the competent non-expert cluster comprised those individuals who were slower, less accurate, less discriminating, and more sequential in the acquisition of information. Consequently, the groups differed in their diagnostic outcomes, with members of the expert cluster more likely to select the optimal diagnostic response, lending support to the view that the groups genuinely reflected distinct levels of diagnostic performance.

The present sample was restricted to qualified practitioners, ensuring that members of both groups possessed sufficient experience to acquire expertise (Simon & Chase, 1973). As expected, a sizeable proportion of the sample failed to meet a level of performance that could be described as expert. This supports the view that expertise does not develop as a linear function of experience, but rather, develops with the capacity to extract and utilize agnostic cues in the environment (Cellier et al., 1997; Lipshitz et al., 2001; Wiggins & O’Hare, 2003). This is also supported by the weak correlations between years of experience in the domain and performance on each of the EXPERTise tasks. However, this analysis should be interpreted with caution due to the restricted range of experience within the sample and the general measure of experience.

The third hypothesis that the competent non-expert cluster would perform at a significantly lower level than the experts when static, but not dynamic, stimuli were available was also supported. While the performance of the competent cluster appeared to diminish in
the absence of dynamic information, the members of the expert cluster performed equally well with both static and dynamic stimuli.

Both dynamic and static stimuli were included to replicate distinct real-world conditions. Although medical practitioners in the pediatric intensive care unit will usually have live updates of a patient’s vital signs, in some circumstances, the bedside monitoring equipment may not receive the necessary trace feeds, in which case, the vital signs are only updated sporadically. However, dynamic (or live) information can provide a wealth of unique cues that are generally regarded as being very diagnostic, including the rate, direction and type of change (Tversky et al., 2002). These cues are not available under static conditions and this presumably impacts diagnostic performance (Shanteau, 1992). Notably, the present results suggest that experts are able to maintain performance in the absence of dynamic information, whereas experienced, but only competent practitioners are less able.

This is not the first study to observe that the distinction in performance between non-experts and experts is greater using static stimuli. For example, Lowe (2001) observed that non-experts produced very simple and inaccurate forecasts compared to domain experts when asked to predict weather patterns based on static meteorological maps.

In terms of expert assessment, it appears that static scenarios provide the greatest distinction between competence and expertise. Therefore, future cue-based assessments of expertise within experienced samples may benefit from using static stimuli, from which patient trends can only be inferred.

Implications and limitations

By distinguishing non-experts and experts within an experienced sample, the present results support the use of EXPERTise as a method of monitoring the performance and skill acquisition of experienced practitioners. By expanding the range and number of practitioners within the EXPERTise database, it will be possible to develop standardized norms of
performance. These norms can be used to determine whether practitioners are developing expertise at a rate that is typical of their experience, position or training, in comparison to their peers. Further, because EXPERTise measures four independent facets of expert performance, it may be used to identify the specific component skills of cue-based diagnosis that competent non-experts are yet to acquire. This can inform specific cue-based remedial training strategies, which have already achieved successes in aviation (Wiggins & O’Hare, 2003), mining (Blignaut, 1979) and general medicine (Litchfield et al., 2010).

It is, however, necessary to address one of the limitations of the present study before EXPERTise can be used as a normative assessment of performance. Specifically, the present study only demonstrated that there are two levels of performance within a sample of experienced practitioners. It did not attempt demonstrate that there is a relationship between the EXPERTise and on-the-job performance. Therefore, there is an opportunity to investigate the relationship between expertise and objective, on-the-job measures of performance, albeit in other domains.

A second limitation of the present study is that it only considered the relationship between performance and the presence and absence of dynamic information. However, dynamism is a more nuanced facet of the operational environment than operationalized here. For example, the present study did not consider whether dynamism might also act as a distracter. In the present study, dynamism may have facilitated performance by drawing attention to relevant diagnostic cues through visual changes, such as parameters ‘flashing’ (Beck, Lohrenz, & Trafton, 2010). However, in some circumstances, the most relevant patient parameter may not be that highlighted by the bedside monitor. Therefore, dynamism-generated salience may, in some cases, impair performance. Future research might compare competent and expert recall for patient parameters that are paired with a parameter in its alert sequence.
Conclusion

The present study was designed to determine whether the application of four independent cue-based assessments could, collectively, distinguish competent from expert practitioners within an experienced sample of pediatric intensive care unit staff. Overall, performance on all four-assessment tasks successfully differentiated the two groups, whereby experienced staff could be divided into competent and expert practitioners based on their capacity for cue utilization.

This study also examined the relationship between performance and dynamic information. Consistent with predictions, competent non-experts performed nearly as well as genuine experts when dynamic information was made available, but less well when only static features were available. This suggests that competent individuals may be less able to reason about dynamic systems based on static sources of information. However, additional research is required to determine which aspects of dynamism facilitate competent performance in non-experts, and to identify potential remedial training targets to improve performance when patient vitals are only sporadically updated.
Key points

- The transition from competence to expertise occurs with the refinement of associations in memory between features of the environment and subsequent events.
- The present study distinguished competent from expert practitioners within an experienced sample using tasks in which the extraction of static and dynamic cues was advantageous.
- The results were most pronounced using static, rather than dynamic, stimuli.
- These tasks can be used to assess diagnostic expertise and guide cue-based training interventions.
References


Tables and figures

Figure 1. Example patient bedside monitor output.
Table 1. Participant cluster means and rank on each task

<table>
<thead>
<tr>
<th></th>
<th>Cluster 1: Competent non-experts (n = 24)</th>
<th>Cluster 2: Experts (n = 26)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Cluster rank</td>
</tr>
<tr>
<td>Static FID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction Time (seconds)</td>
<td>11.08 (4.36)</td>
<td>2</td>
</tr>
<tr>
<td>Dynamic FID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction Time (seconds)</td>
<td>4.80 (1.50)</td>
<td>2</td>
</tr>
<tr>
<td>Static FID Recall Accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.29 (2.05)</td>
<td>2</td>
</tr>
<tr>
<td>Dynamic FID Recall Accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.96 (1.80)</td>
<td>2</td>
</tr>
<tr>
<td>Sequential PAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction Time (seconds)</td>
<td>5.98 (2.16)</td>
<td>2</td>
</tr>
<tr>
<td>Sequential PAT Response Variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.52 (.62)</td>
<td>2</td>
</tr>
<tr>
<td>Simultaneous PAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction Time (seconds)</td>
<td>4.28 (1.67)</td>
<td>2</td>
</tr>
<tr>
<td>Simultaneous PAT Response Variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.22 (.68)</td>
<td>2</td>
</tr>
<tr>
<td>Static FDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Variance</td>
<td>2.73 (2.85)</td>
<td>2</td>
</tr>
<tr>
<td>Dynamic FDT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Variance</td>
<td>4.13 (4.37)</td>
<td>2</td>
</tr>
<tr>
<td>Static TT Proportion (ratio)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.91 (.17)</td>
<td>2</td>
</tr>
<tr>
<td>Dynamic TT Proportion (ratio)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.92 (.09)</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 2. Frequency count of optimal and suboptimal responses to the static feature discrimination task by cluster

Figure 3. Frequency count of optimal and suboptimal responses to the static transition task by cluster
Appendix A
Static Feature Discrimination Task Scenario

Scenario

2-month old baby girl with increasing coughing and vomiting feeds for last 2 days. Admitted to HDU. NPA viral screen is negative. Chest X-ray shows bilateral hyperinflation. She has been managed by withholding feeds and by giving IV fluids at normal maintenance, and nasal prong oxygen. Four hours after admission to HDU, she is noted to be increasingly restless with marked intercostal recession and grunting between frequent episodes of coughing.

Response Options

- Continue with 2L/min nasal prong oxygen
- Give sedative medication
- Give nebulized adrenaline and IV dexamethasone
• Give a hypertonic saline nebulizer
• Collect a capillary blood gas and act depending on the measured CO2 level
• Assist ventilation with CPAP
• Intubate urgently and ventilate via an endotracheal tube
• Do nothing

**Ratings of Feature Utility**

• Patient history
• Neurological observations
• Respiratory observations
• Heart rate
• Respiratory rate
• SaO2
• Blood pressure
• Patient Age
• Chest x-ray
Appendix B

Static Transition Task Scenario

Scenario

You are asked to assess a 4-year old boy ventilated in PICU with low urine output.

Feature: Details

- Admission diagnosis: Neutropenia sepsis
- History prior to admission: 24-hours increasingly unwell with abdominal pain, diarrhea, fever, lethargy
- Previous medical history: recent diagnosis of acute lymphoblastic leukemia
- Current drug therapy: IVIs morphine@20mcg/kg/hr, midazolam@1mcg/kg/min, dopamine@5mcg/kg/min; IV timentin, gentamicin, omeprazole
- Level of parental anxiety: High
- Peripheral perfusion: Hands and feet are warm
• Capillary refill time: 4 seconds
• Peripheral pulse character: Easily palpable
• Arterial blood lactate: 4.5 mmol/L
• Central venous saturation: 32%
• Level of sedation: Responds only to painful stimulus
• Vital sign trend: Heart rate increasing, BO maintained with wide pulse pressure
• Abdominal examination: Soft, liver 4cm below costal margin, catheter in sit, no bladder palpable
• Skin examination: Widespread erythema
• Chest auscultation: Good bilateral air entry and normal breathe sounds
• Cardiac auscultation: Heart sounds normal, no murmur
• Pupillary reaction to light: Pupils R. = L., 2-3mm and equally reactive to light
• Blood urea and electrolytes: Na 132mmol/L, K 4.0 mmol/L, urea 6.2mmol/L, creatinine 52umol/L
• Full blood count: Hb 9.5g/dL, WCC 0.1, Plats 52
• Ventilator settings: Rate 25/min, FiO2 0.60, PIP 25 cmH2O, PEEP 10 cm H2O, Ti 1.0s

Response Options

• Increase the FiO2 to 100%
• Start a norepinephrine (noradrenaline) infusion
• Give 1mcg/kg IV hydrocortisone
• Give a 3% hypertonic saline bolus
• Give a packed cell transfusion
• Flush the urinary catheter
• Give an IV isotonic fluid bolus
• Repeat the serum urea and electrolytes in 4-hours
## Appendix C

Paired Association Task Items

<table>
<thead>
<tr>
<th>Sequential Pairs</th>
<th>Simultaneous Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Feature A</strong></td>
<td><strong>Feature B</strong></td>
</tr>
<tr>
<td>Breathing</td>
<td>Sedation</td>
</tr>
<tr>
<td>Cardiac catheter</td>
<td>Coagulation</td>
</tr>
<tr>
<td>Bleeding</td>
<td>Anesthetic</td>
</tr>
<tr>
<td>Induction</td>
<td>Problem with drug</td>
</tr>
<tr>
<td>Low feet temp</td>
<td>Low feet pulsation</td>
</tr>
<tr>
<td>Echo</td>
<td>Neurological</td>
</tr>
<tr>
<td>Arrythmias</td>
<td>Low C02</td>
</tr>
<tr>
<td>Hypotension</td>
<td>Problem with drug</td>
</tr>
<tr>
<td>Mucus in ventilation</td>
<td>Desaturation</td>
</tr>
<tr>
<td>Breathing problems</td>
<td>Oxygenation</td>
</tr>
<tr>
<td>Inhibition of diaphragm</td>
<td>Desaturation</td>
</tr>
<tr>
<td>Bleeding</td>
<td>Circulation</td>
</tr>
<tr>
<td>Extreme tachycardia</td>
<td>Ventilation issues</td>
</tr>
<tr>
<td>Temperature</td>
<td>Heart rhythm</td>
</tr>
<tr>
<td>Magnesium</td>
<td>Bleeding in abdomen</td>
</tr>
<tr>
<td>Low heart rate</td>
<td>Ventilation obstruction</td>
</tr>
<tr>
<td>Lactate acid</td>
<td>Organ perfusion</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>Pericardium</td>
</tr>
<tr>
<td>Fast arrhythmia</td>
<td>Cardiac arrest</td>
</tr>
<tr>
<td>Ventricular function</td>
<td>Issues with perfusion</td>
</tr>
<tr>
<td>Low blood pressure</td>
<td>Ventricular dysfunction</td>
</tr>
<tr>
<td>High potassium</td>
<td>Issues with perfusion</td>
</tr>
<tr>
<td>High lactate</td>
<td>Liver issues</td>
</tr>
<tr>
<td>Long bypass time</td>
<td>Heart failure</td>
</tr>
<tr>
<td>Excess bleeding</td>
<td>Cardiac injury</td>
</tr>
<tr>
<td>Chest drain volume</td>
<td>Ventilation</td>
</tr>
<tr>
<td>Pulmonary vascular restriction</td>
<td>Low C02</td>
</tr>
<tr>
<td>Cardiac surgery</td>
<td>Low blood pressure</td>
</tr>
</tbody>
</table>