Role and experience determine decision support interface requirements in a neonatal intensive care environment

Gary Ewing, a,* Yvonne Freer, b Robert Logie, c Jim Hunter, a Neil McIntosh, b Sue Rudkin, c and Lindsey Ferguson b

a Department of Computing Science, King’s College, University of Aberdeen, Aberdeen, UK
b Neonatal Intensive Care Unit, Simpson Centre for Reproductive Health, New Royal Infirmary, Edinburgh, UK
c Department of Psychology, King’s College, University of Aberdeen, Aberdeen, UK

Received 20 March 2003

Abstract

The aim of this paper is to describe a novel approach to the analysis of data obtained from card-sorting experiments. These experiments were performed as a part of the initial phase of a project, called NEONATE. One of the aims of the project is to develop decision support tools for the neonatal intensive care environment. Physical card-sorts were performed using clinical “action” and patient “descriptor” words. Thirty-two staff (eight junior nurses, eight senior nurses, eight junior doctors, and eight senior doctors) participated in the actions card-sorts and the same number of staff participated in separate descriptors card-sorting experiments. To check for consistency, the card-sorts were replicated for nurses during the action card-sorts. The card-sort data were analysed using hierarchical cluster analysis to produce tree-diagrams or dendrograms. Differences were shown in the way various classes of staff with different levels of experience mentally map clinical concepts. Clinical actions were grouped more loosely by nurses and by those with less experience, with a polarisation between senior doctors and junior nurses. Descriptors were classed more definitively and similarly by nurses and senior doctors but in a less structured way and quite differently by junior doctors. This paper presents a summary of the differences in the card-sort data for the various staff categories. It is shown that concepts are used differently by various staff groups in a neonatal unit and that this may diminish the effectiveness of computerised decision aids unless it is explored during their development.

2003 Elsevier Inc. All rights reserved.

Keywords: Card-sorts; Concept-sorts; Cluster analysis; User-interface; Neonatal intensive care; Clinical terminology; Clinical roles

1. Introduction

The modern intensive care unit is an environment that requires medical and nursing staff to deal with large amounts of, and many different types of, information in making clinical decisions. It has been shown that just displaying these data in their raw form does not of itself lead to improved patient care [1]. The work reported in this paper formed part of the initial effort in an ongoing project, NEONATE, to develop decision support for clinical staff (doctors and nurses) in a neonatal intensive care environment—The Neonatal Intensive Care Unit (NICU) at the Simpson Centre for Reproductive Health, Edinburgh. The component of this initial phase, which is reported here, focused on developing a concise lexicon of terms used by clinical staff during clinical practice. This information was used to design the user-interface for a software tool to allow a trained observer (research nurse) to record (in a standardised fashion on a computer database) the clinical activities of doctors and nurses as a data gathering exercise for NEONATE. Detailed physiological data such as heart rate, blood pressure, were automatically collected by the NICU’s computerised monitoring system.

These lexicons were further used as the basis for concept-sorting experiments designed to elucidate the way clinical staff mentally organise those terms. Each term (concept) in a specific lexicon was transcribed on to

* Corresponding author. Fax: +44-1224-27-3422.
E-mail address: gewing@csd.abdn.ac.uk (G. Ewing).
a card, and individual subjects were invited to sort the cards into piles, such that all of the concepts in a given pile were related in some way—as determined by the subject. The interest in such an experiment is to see the extent to which all subjects (or all members of a specific subset of subjects—e.g., junior nurses) consistently group certain concepts together.

One method of analysing such card-sort experiments is by performing cluster analysis in order to generate tree-diagrams (or dendrograms) as a graphical representation of the relationships between the concepts under study. A cluster exists when two or more concepts are grouped together (deemed similar). An introduction to cluster analysis for concept-sorting experiments is given in this paper.

Using these tools, this study assesses how nursing and medical staff use language and concepts in a neonatal unit. We speculated that within any unit staff would possess different degrees of knowledge and experience, which would result in different needs (in terms of user-interfaces for clinical decision support). Words and concepts might be used differently because of these differences. Our results indicate that the design of a NICU-based decision support tool needs to consider the different perspectives of the various staff involved.

2. Methodology

The scope of this section is to outline the methodology employed in the carrying out and the analysis of the card-sorting experiments.

NICU staff were classified as senior nurses, senior doctors, junior nurses, and junior doctors, in order to delineate their roles in the unit and the vocabularies they use to categorise the neonatal data they obtain by observation and physical means. The nurses were categorised as senior or junior depending on their years of working in a neonatal unit. The medical staff were categorised by their experience working in a neonatal unit. The level at which staff were categorised, whether junior or senior, was decided upon by the senior clinical staff involved in the project.

Staff classifications were determined as follows. Junior nurses are 'general' or 'sick children' trained nurses and/or midwives with or without a specialist qualification in neonatal intensive care. Equally, they may or may not be degree level educated. Junior nurses are differentiated from senior nurses in that they do not take charge, and are not responsible for the day to day running and management of the NICU. Senior doctors are consultant level or those considered to be appropriately trained and experienced to accept consultant level responsibility. We do not know how these definitions compare internationally.

2.1. Clinical lexicons

In order to elicit lexicons for both patient “descriptors” and clinical “actions” a research psychologist interviewed medical and nursing staff with various levels of experience. Four hundred and nineteen actions and 520 descriptors were offered by participating staff. Senior medical and nursing staff subsequently reviewed these lists for consistency, and to remove synonyms and singletons (single words used by only one member of staff). The derived actions lexicon contains 51 terms, while the descriptors lexicon contains 166 terms (see Section 4.1 for details).

2.2. Card-sorting experiments

Following the interviews, we carried out card-sorting experiments using the two lexicons containing the concepts to be sorted. Card-sorting was used as an elicitation technique because “Concept Sorting” is well-known, and studies in Cognitive Psychology and related fields [2,3] have shown it to be effective and very efficient.

Thirty-two subjects consisting of eight junior nurses, eight senior nurses, eight junior doctors, and eight senior doctors, participated in the actions card-sorts and 32 staff (of the same levels) participated in separate descriptors card-sorting experiments. There was considerable overlap of staff in the two groups (of 32), but the actions and descriptors experiments were held several months apart. At least one week after the initial sessions, the card-sorts were repeated, for the nurses only, in the actions experiments (because of practical constraints). Each actions card-sorting session took about an hour to complete on average, while each descriptors card-sort session required about 1.5 h to complete.

The actual card-sorting procedure that we asked subjects to perform is illustrated in Fig. 1. During a session, each subject was presented with a physical pile of cards, with the front of each card containing a term from the appropriate lexicon. These cards were marked on the back with a bar-code (3 of 9 code), containing the term on the front of the card, and a unique identifying alphanumeric code.

Each subject was asked to sort the cards into piles of “similar” cards, without any prompting as to how many piles to create or what attributes to use to sort the cards. The experimenter then entered the names and codes of the cards, within their sorted groups, into a

---

1 Many of the staff who participated in the card-sorts were also involved in the interviews to establish the lexicons. However, there was several months gap between the interviews and the card-sorts.

2 Subjects were actually asked to sort concepts into hierarchical groups, but we do not use the information on higher (than first) level groupings in this paper.
computer database by means of a bar-code scanner. This saved considerable time and minimised errors in data entry.

2.3. Analysis of the card-sort experiments

Within the context of the work described in this paper, cluster analysis calculates the distance of the perceived relationships between concepts, and displays these relationships graphically (dendrograms). Thus, if concepts (cards) tend to be put in the same pile by most subjects, the distance between them is small. However, if concepts tend to be sorted into different piles the distance is large.

We performed hierarchical agglomerative cluster analysis of the concept-sort data using a free software package called EZCalc [4], which was designed to be used with its companion software, EZSort, that facilitates computer-based concept-sorting experiments. As we did physical sorts (though semi-automated), the data files needed to be formatted and pre-processed to allow use by EZCalc. The pre-processing also checked for consistency within the concept-sort data files. We employed the average linkage method of cluster analysis, as this provides a good compromise between the extremes of other methods [5] (see Section 3.3 for a description of linkage methods).

Using in-house software, we carried out further analysis of the concept-sort data. This software produced distance matrices, which quantified how often all the possible pairs of concepts were grouped together by each class of subjects. More specifically, a similarity matrix was produced from the frequencies that pairs of concepts appeared in the same pile, which were then normalised by the number of sorters who put those concepts together; e.g., if two concepts were always put in the same pile their distance value would be 0, or if they were put in the same pile by say 80% of the subjects the distance value would be 0.2. This measure of distance is called “percent disagreement.”

3. Development of dendrograms from concept-sort data

In this section, we discuss in detail the distance matrix and give an outline of the clustering process used to produce the dendrograms. A full discussion is beyond the scope of this paper, but is available elsewhere (e.g., [5–8]).

3.1. The distance matrix

As described in Section 2.3, one of the products of a card-sorting experiment is a distance matrix, where the elements have values between zero and unity. Since distance matrices are symmetrical about the major diagonal, we converted them to triangular matrices thus:

\[
D = \begin{bmatrix}
    d_{11} & d_{12} & \cdots & d_{1N} \\
    0 & d_{22} & \cdots & \vdots \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & d_{NN}
\end{bmatrix}
\]

(1)

Note, each of the elements in the main diagonal represents the distance between a concept with itself, and are not used in clustering. Each of the elements \(d_{pq}\) of the distance matrix \(D\) has a value within the range zero to one, and represents (1-the joint probability \(P(x_p, x_q)\); i.e., \(P(x_p, x_q)\) is the (a posteriori) probability that the two concepts \(x_p\) and \(x_q\) appear together in the same pile in a concept-sort. The distance matrix, \(D\), contains \(N(N-1)/2\) elements (not including the main diagonal), the number of possible combinations of different pairs of cards from a stack of \(N\) cards.

However, since all of the distance values in the matrix do not sum to one; i.e.

\[
\sum_{p=1}^{N} \sum_{q=1}^{N} d_{pq} \neq 1, \quad (2)
\]

the elements of the distance matrix as a whole do not represent a probability distribution.

3.2. Agglomerative clustering

In Agglomerative Clustering, each concept is initially placed in its own group. Therefore, if we have \(N\) concepts, the minimum number of steps required to form a single cluster is \(N\), and the maximum is \(N(N-1)/2\). When datasets are large, the calculation of distance matrices and the computation of agglomerative clustering can be time-consuming. The cluster analysis requires that each concept is initially placed in its own group.
concepts to cluster, we start with \( N \) groups. Each of these groups contains a single concept only; i.e., each group is initially a ‘cluster’ of one. To run an agglomerative clustering, you need to decide upon a method of measuring the distance between two concepts and apply a strategy on how to choose potential members of each cluster. In the case of the former, the distance measure we used to develop the distance matrices was “percent disagreement” and was described in Section 2.3. In the case of the latter, we chose a clustering strategy called average linkage. Some clustering, or linkage, strategies are discussed shortly in Section 3.3.

The simplest way to cluster the concepts is to find pairs with minimum distance; i.e., find the pair of concepts with the minimum distance and cluster, then find the next pair with minimum distance and cluster, known as simple linkage. Note that pairs may be made with both the original concepts and the generated clusters. However, this would produce a dendrogram where each successive cluster would be included in the next level cluster. In the case of the former, the distance measure was “percent disagreement” and was described in Section 2.3. In the case of the latter, we chose a clustering strategy called average linkage. Some clustering, or linkage, strategies are discussed shortly in Section 3.3.

The simplest way to cluster the concepts is to find pairs with minimum distance; i.e., find the pair of concepts with the minimum distance and cluster, then find the next pair with minimum distance and cluster, known as simple linkage. Note that pairs may be made with both the original concepts and the generated clusters. However, this would produce a dendrogram where each successive cluster would be included in the next level cluster, and each cluster would have only two members (unless more than two are exactly the same distance apart). This would artificially restrict the dendrogram.

A way to overcome this difficulty is to use different linkage methods to that of purely simple linkage. Two other options are average linkage and complete linkage.\(^6\)

### 3.3. Linkage methods

Descriptions of three commonly used methods for finding linkages during clustering are given below.

#### 3.3.1. Simple linkage (nearest neighbour)

The nearest neighbour method evaluates the distance between two clusters as the shortest distance that can be found between any pair of members of the two clusters under consideration. This tends to produce trees with long “chains.”

#### 3.3.2. Average linkage

The average linkage method evaluates the distance between two clusters as the shortest average distance that can be found between members of the two clusters under consideration. This can cause the groups to have similar diameters in each metric direction, though each group can be a different size. This was the method we employed.

#### 3.3.3. Complete linkage (furthest neighbour)

The furthest neighbour method evaluates the distance between two clusters as the longest distance that can be found between any pair of members of the two clusters under consideration. This method performs well, but may cause the clustering of eccentric or prolonged groups.

In the next section the results of the card-sorting experiments for both action and descriptor words are presented.

### 3.4. Interpretation of dendrograms

As alluded to earlier, dendrograms give a graphical representation of how a group of subjects sorted cards, representing concepts, into piles of ‘similar’ concepts, based on some subjective notion of similarity. The aim is to produce a diagram representing how a group of subjects “mentally map” the concepts under study. The dendrogram is derived by clustering the distance matrix as explained earlier in this section.

The \( y \)-axis of the dendrogram lists the concepts that were sorted (e.g., actions), while the \( x \)-axis represents the distance between the concepts within a cluster, which is represented by a vertical line joining the concepts within a cluster.

Consider Fig. 3. An example of a cluster is the group containing the concepts (actions here) of ‘containment,’ ‘comfort,’ and ‘cuddle-kangaroo care,’ which seem to have been put in the same card pile by all of the junior nurses, giving a distance of 0.0. As you move (right) along the \( x \)-axis of the dendrogram, the ‘distance’ between actions within each cluster gets larger, meaning that fewer junior nurses put those actions together. As another example, ‘feeding’ and ‘comfort’ have a ‘distance’ of about 0.27, meaning that a majority of junior nurses put these two in the same pile of cards, but some of them allocated each of these two actions to two different piles of cards. Also, as you move right along the \( x \)-axis of the dendrogram, tight clusters are progressively grouped into larger loose clusters. For example, the cluster consisting of ‘communication’ and ‘parent education’ (clustered at about 0.13) is clustered with ‘reading’ at about a distance of 0.32.

### 4. Results

The results of the cluster analysis are presented first for actions and second for descriptors within each subsection.

#### 4.1. Development of lexicons

When developing the “actions” lexicon, senior nurses gave 134 separate action expressions, junior nurses 99, senior doctors 75, and junior doctors 108—a total of 416 but by eradicating duplicates between groups this was reduced to 193 different actions. Similarly when developing the “descriptor” lexicon, senior nurses gave 520 separate descriptor terms, junior nurses 258, senior
doctors 461, and junior doctors 358—a total of 1796 which was reduced to (coincidentally) 520 by eradicating duplicates. The discarding of terms used by only a single individual and the amalgamation of the synonyms reduced the total number of different actions to 51 and the total number of different descriptors to 166.

The results of the lexicon development were used to design the user-interface of a data collection tool (called BabyWatch [9]) used by a research nurse to record observational data on patients and clinical interventions. A page of the user-interface for BabyWatch is shown in Fig. 2 displaying the actions lexicon. The large set of descriptors was divided into seven major groups (by senior medical and nursing staff)—feeding, crying, sleep, movement-muscle tone, skin, size-weight-shape, and bowel-urine. Shown in Table 1 are examples of the descriptor terms for the “movement-muscle tone” subset and synonyms that were elicited at interview.

4.2. Consistency of the concept-sorting results

As indicated, in Section 2.2 the actions concept-sorts were replicated for (the same groups of) nurses in the experiments.

The initial and replicated distance matrices were converted into vectors by concatenating the rows of the distance matrices into one column per matrix of distance values.

These vectors were statistically compared using the Wilcoxon signed ranks test (as a non-parametric alternative to the t-test, since plots of the data indicated they were not normally distributed) [10], and by the performance of a Pearson correlation analysis.

The correlation analysis yielded the following results: junior nurses \( r = 0.91, p < 0.0001 \) and senior nurses \( r = 0.92, p < 0.0001 \). These results indicate that there exists a very high degree of correlation between the results for the first and second card-sorts for both junior and senior nurses. In other words, the results of the first and second card-sorts are consistent. However, the results of the Wilcoxon test \( (p < 0.00001, \text{ for both junior and senior nurse data}) \) indicate that the initial and replicate data come from different probability distributions (more specifically the distributions have different medians). Note that the large sample sizes would have made this test very sensitive to any differences.

The statistical analysis seems to indicate that the overall results for the first and second card-sorts are similar, but there are differences on a more detailed scale. This may reflect differences within a staff class (i.e., individual differences within a group) that are expressed slightly differently on the occasion of the replicated card-sort compared to the initial card-sort. That is, there may be some instability within a group. This has not been specifically addressed in this paper, as no measurements of consistency among the various members of a given staff class.
group are presented. It is possible that apparent differences between groups simply represent inhomogeneity with some groups. It is worth noting that if the underlying distributions themselves (i.e., from a group) are unstable because members of the group are not consistent among themselves, the validity of intergroup comparisons is reduced.

The statistical analyses were performed using SPSS version 11.0 for Windows.

### 4.3. Dendrograms

As described in Section 2.3, cluster analysis was performed on the processed card-sorting data and dendrograms were produced for the various classes of staff. The dendrograms graphically illustrate how doctors and nurses mentally envisage the actions and descriptors within a conceptual framework. Therefore, it is appropriate to compare the conceptualisation of descriptors and actions of the different staff classes because the intention is to see how applicable a generic computer-assisted decision-making tool would be. Further, we expected that the data and dendrograms would be different (because of the nature of the professional roles/experience), but we did not know how different. This was the aim of this study. This information is important because it quantifies the relationship between simple concepts applicable to neonatal care and therefore helps draw conclusions as how to proceed with the development of the decision tools.

This analysis has yielded some interesting results, which are briefly discussed here.

#### 4.3.1. Dendrograms—actions

Fig. 3 shows a section of the dendrogram derived from the actions card sorts for junior nurses, while Fig. 4 displays a section of the dendrogram derived from the actions card sorts for senior doctors. Both of these sections of dendrograms show some common terms of interest from the actions lexicon.

In appearance, the clustering of action cards appeared more variable than descriptor cards (see Section 4.3.2). We define a cluster as a grouping together of two or more concepts, in the dendrograms, at a given

---

Table 1

<table>
<thead>
<tr>
<th>Movement/muscle tone</th>
<th>Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floppy</td>
<td>Hypotonic, poor tone, flaccid, limp, flat, poor movement, weak movement</td>
</tr>
<tr>
<td>Hypertonic</td>
<td>Stiff, rigid, poor tone, tense, increased tone</td>
</tr>
<tr>
<td>Jittery</td>
<td>Jerky, twitchy, tremulous</td>
</tr>
<tr>
<td>Active</td>
<td>Vigorous</td>
</tr>
<tr>
<td>Irritable</td>
<td>Agitated, restless, jumpy</td>
</tr>
<tr>
<td>Lethargic</td>
<td>Sleepy, drowsy, inert, inactive, not moving, lying still</td>
</tr>
<tr>
<td>Good tone</td>
<td>Normal tone, handles well, movement appropriate for gestational age, normal movement, appropriate response, handling well tolerated, symmetrical movement, relaxed, normal posture</td>
</tr>
<tr>
<td>Responsive</td>
<td>Active on handling</td>
</tr>
<tr>
<td>Wriggly</td>
<td>Squirming</td>
</tr>
<tr>
<td>Unresponsive</td>
<td>Non-reactive, not responding</td>
</tr>
<tr>
<td>Fits</td>
<td>Convulsions, tonic, clonic, cycling movement</td>
</tr>
<tr>
<td>Back arching</td>
<td>Opisthotonic</td>
</tr>
</tbody>
</table>

Column 2 shows synonyms that were yielded during lexicon elicitation interviews.

---

Fig. 3. Section of the “Actions” dendrogram for junior nurses.
distance (along a vertical line from a specified distance on the x-axis).

We considered cases where there was 100% agreement within a group (i.e., single actions or clusters containing actions with zero distance between them). We counted the total number of clusters of two or more single actions for each staff group and these are shown in Table 2. To show exactly how the clusters were counted a breakdown is given for senior nurses. The clustering for senior nurses included 4 pairs, two groups of 4 actions, and one group of 5 actions, which totalled to 21 actions, or cards out of 51 (in the actual sort). Table 2 gives a summary of how the card-sort data from the four different staff groups were clustered at the 100% agreement level. Of particular interest, are the numbers in the columns 'Average Cluster Size' and 'Percentage of Actions Clustered.' The former provides a crude measure of the level of discrimination that a group employed in placing cards in the same pile, while the latter gives an indication of the level of agreement within a group in placing cards in the same piles. A full analysis of discrimination versus agreement is beyond the scope of this paper. However, from the table, we can see that doctors tended to generate smaller card piles than nurses, implying that doctors applied more discrimination, but they only completely agreed on the sorting of 27% or less of the actions cards compared to greater than 41% by nurses.

The only cluster that was common to all groups and all participants contained the actions of “giving incubator oxygen” and “giving nasal oxygen.” The two nursing groups showed commonality on five occasions in clustering items such as intravenous fluid management, oxygen delivery, gavage feeding, comfort strategies, and skin care. The two doctor groups showed similarity in two clusters: delivery of oxygen and skin care.

There is a difference in structure of the dendograms across the different staff groups, and an extreme difference in structure between junior nurses and senior doctors. In the case of the latter, there is evidently a much richer structure with more groups (more discerning) than is evident for the former, who formed large groupings with little discernment. For example, junior nurses did not tend to place the “Biophysical Observation” card with the “Examine Baby” card (whereas the other three staff groups [senior nurses, junior doctors, and senior doctors] grouped these closely together) nor do they seem to be helped much by the computer display of physiological data—“Observe Baby” was not grouped with “Biophysical Observation.”

It is clear from the dendrograms that the various groups of staff within the NICU interpret and categorise

![Fig. 4. Section of the “Actions” dendrogram for senior doctors.](image)

Table 2
The clusters of actions at 100% agreement for the four staff groups, which is derived from the dendrograms

<table>
<thead>
<tr>
<th>Staff class</th>
<th>Number of clusters</th>
<th>Number of actions clustered</th>
<th>Average cluster size</th>
<th>Percentage of actions clustered (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior nurses</td>
<td>10</td>
<td>23</td>
<td>2.4</td>
<td>47</td>
</tr>
<tr>
<td>Senior nurses</td>
<td>7</td>
<td>21</td>
<td>3.0</td>
<td>41</td>
</tr>
<tr>
<td>Junior doctors</td>
<td>6</td>
<td>13</td>
<td>2.2</td>
<td>25</td>
</tr>
<tr>
<td>Senior doctors</td>
<td>7</td>
<td>14</td>
<td>2.0</td>
<td>27</td>
</tr>
</tbody>
</table>

Note 51 actions were sorted.
data differently. The grouping of data appeared to be associated with particular professional practice. For example the actions related to artificial ventilation were, in general, grouped in the same way for senior and junior nurses; senior doctors had a similar group but omitted the management of the ventilator humidifier—this was grouped with issues of equipment safety. Like senior doctors, junior doctors clustered actions related to artificial ventilation but in their minds, humidifier management and equipment safety were associated with routine nursing care. These variations in representation may correspond to differences in knowledge and or professional role and responsibilities.

### 4.3.2. Dendrograms—descriptors

Displayed in Fig. 5 is a section of the dendrogram derived from the descriptors card-sorts for junior nurses, while Fig. 6 displays a section of the dendrogram derived from the descriptors card-sorts for junior doctors. Counting clusters containing descriptors with zero distance between them, we obtained the results displayed in Table 3, which as was shown for Table 2, gives a summary of how the card-sort data from the four different staff groups was clustered at the 100% agreement level. Table 3, compared to Table 2, indicates that the differences in the card-sorting strategies used by the four staff groups were considerably less pronounced during the descriptors card-sorts than the actions card-sorts.

It is interesting to note that on no occasion did all staff groups cluster the same descriptors, together into a single entity. However, on 57 occasions, two or more descriptors were clustered together by two or more staff groups. For example: “skin perfusion,” “shutdown,” and “poor capillary return” were always clustered together by all staff, but in addition junior doctors included “mottled” and “poor colour,” and junior nurses included “blue,” “mottled,” and “poor colour.”
Other similarities were associated with describing stools, urine, sleep wake state, levels of consciousness, and skin quality.

5. Discussion

The purpose of carrying out the card-sorts was to find out how different groups of people conceptualise knowledge. We are able to show that groups of staff with different roles and experience conceptualise their knowledge in different ways. The lower within-group consistency (compared to the other three groups) implied by the ‘descriptor’ card-sort results for the junior doctors may reflect a lack of knowledge. However, many of the terms used are specific to newborn care, and in particular to the unit in which the study was conducted. It is more likely that the difference seen in the junior doctors’ clustering is an indication of the short time that they had spent on the unit and of their therefore not having had the opportunity to acquire the vernacular. The action cards create a more complex picture. There is less agreement between the professional groups and it may be that the differences are a reflection of the nature of the actions themselves. On the whole, the actions represent nursing activities. The way these are organised depend not only on their relationships to one another but also to expertise, the theory or model that the individual practitioner uses in cue acquisition and competing goals when combining actions [11–13]. Knowing that staff conceptualise knowledge in different ways raises a number of questions.

- The decision-making process has been classified by both the construction of data and complexity of the task itself [14]. Therefore, how do the differences seen in the current study in clustering information impact on decision making?
- Where there is blurring of boundaries in professional role and function, what are the consequences of the differences seen in the current study on decision-making processes and outcome?
- In specialised areas where there is considerable data acquisition and influence on reasoning and decision making, as in a neonatal intensive care unit, how may the differences seen in the current study influence the development of computer-assisted decision tools?

Offredy [15] has shown that although processes in decision making are different between nurse practitioners and general practitioners, outcomes as measured in accuracy of diagnosis, are similar. This situation is, however, different from that within the neonatal unit as not all staff within the unit function at the expert level and practitioners would not necessarily be confronted with multiple and complex information that is readily available in the NICU. Computerised systems are seen as one way in which information can be assimilated to assist decision making. It is argued, however, that utilisation of such a system is dependent upon the interrelation of the system’s development, implementation, and functioning with the skilled and pragmatically oriented work of various health professionals [16].

6. Summary and conclusion

We have described card-sorting experiments designed to elicit knowledge about how nursing and medical staff in a Neonatal Intensive Care Unit mentally map clinical concepts. These experiments produced data on how the subjects group concepts based on some notion of similarity.

Our results indicate that different staff groups have different needs from a decision support tool on a NICU. We have previously shown that the different staff groups have differing abilities at interpreting trended physiological information [17]. This is possibly due to nurses and doctors having different knowledge bases and roles within health care. Reliance for data interpretation falls heavily on nursing staff that are constantly at the bedside in an intensive care unit, usually with a patient nurse ratio of near 1:1.

We would suggest that the development of decision support tools, particularly in an intensive care environment requires an understanding of the cognitive background of the staff that will use the support. It is crucial that the language and concepts are common between the groups developing and using the tool. These systems are developed with the knowledge of the designer whose knowledge and experience base will be different from the front line staff that are to use them.
Finally, we believe this work contributes to the field of medical/nursing informatics. There is little literature on the analysis of concept-sort data. Further, there is also virtually no literature commenting on the comparison of how nurses and doctors mentally map clinical concepts, particularly in intensive or critical care environments. This paper adds to that scant literature.

Acknowledgments

This work was undertaken as part of the NEONATE project; we are grateful to both the UK ESRC and EPSRC for providing funding under the People at the Centre of Communication and Information Technologies (PACCIT) program. We also acknowledge the input of our co-workers on that project: Peter Badger and Alan Maxwell.

References