HUMAN RELIABILITY AS A SOURCE OF ERROR IN RESEARCH

Agustin Chevez Bernaldo de Quiros[1], Robert Lopez[2], Guillermo Aranda-Mena[3] and Peter Edwards[3]

[1, 3]*School of Property, Construction and Project Management, RMIT University, GPO Box 2476V, Melbourne VIC 300*

[2]*Cooperative Research Centre for Construction Innovation, Department of Construction Management, Curtin University of Technology, GPO Box 1987, Perth, WA 6845, Australia*

Human error is embedded in every human endeavour. Given that research is conducted by humans, it is therefore prone to human error. This paper focuses on the errors that derive from transcribing data across formats. These "input" errors arise largely from the monotony of the data entry process and may mean that an otherwise thoroughly designed research can potentially produce misleading conclusions. The paper reports the results of a quality checking process developed to monitor the transcription of data from paper-based questionnaires, collected as part of current PhD research, into the computer. Following the same entry method, the data from all questionnaires received were input twice by the PhD candidate, then twice again by another participant. The 28,140 entries were matched and any differences analysed in order to quantify the occurrence of input errors committed and identify the nature of these errors. The results suggest that where the input errors were committed had more impact on the findings revealed from each question than the total number of input errors committed.

Keywords: design, error, human reliability, research.

INTRODUCTION

An error essentially involves a deviation of some kind, whether it is a departure from the intended course of actions, departure from a path of actions planned toward a desired goal or deviation from the appropriate behaviour at work (Reason and Hobbs 2003, p.39). When a research project commences, there is usually a general intention to "get it right the first time" (Manavazhi 2004, pp.48-49). However, as individuals make all the decisions regarding what is done, how it is done and who does it, all errors originate from humans. Therefore all errors are ultimately human errors (Kaminetzky 1991, p.5; Sunyoto and Minato 2003, p.298).

This implies that research is prone to the inevitability of human error. In theory this inevitability may be the result of a social cognitive distribution, where everyone has his or her own unique way of thinking despite the teaching and training of others (Busby 2001, p.253). Performing tasks during research will inevitably require cognitive effort from the researcher(s). Cognitive effort involves devoting mental resources to process information and form beliefs (Busby 2001, p.291). Therefore, a basic principle of managing errors is that mistakes can be made by the best of us (Reason and Hobbs 2003, p.16).

[1] s3032925@student.rmit.edu.au

de Quiros, A C B, Lopez, R, Aranda-Mena, G and Edwards, P (2008) Human reliability as a source of error in research. *In:* Dainty, A (Ed) *Procs 24th Annual ARCOM Conference,* 1-3 September 2008, Cardiff, UK, Association of Researchers in Construction Management, 687-696.

This paper will firstly provide background information on the nature of errors committed during research activities and the resultant rework these create. Data entry errors are then investigated using a current PhD research project which focuses on the evolution of workplace architecture as a consequence of technological development in Australia. The investigation focuses on those errors committed whilst manually transcribing data from the paper-based questionnaires returned into a computer software program for data processing. For the purpose of differentiating these with the other errors that may be committed whilst conducting research, these errors will be referred to within this paper as 'input errors'. The objective of the paper is to demonstrate a data entry approach to minimise such errors.

Error and rework in research

Skill- or performance-based error is of particular concern in the research process. This type of error is associated with lapses arising from carelessness or neglect. Errors may involve recognition (misidentifications or non-detection), memory (lapses) or attentive (distractions or slips during action) failures in processing information (Kaminetzky 1991, p.5; Reason and Hobbs 2003, pp.39-41). A concern is that no countermeasure exists for errors in general (Reason 2002, p.40). Attention has previously focussed on human error within systems engineering but with no apparently obvious approach emerging toward the subject (Busby 2001, p.235). To this end, the focus of this paper is on the input errors committed during research.

Importantly, errors are not only caused by deficiencies individual skill and experience, nor are they always the result of isolated mental glitches (Reason and Hobbs 2003, p.10), but are also consequences of surrounding local circumstances and conditions including the task, tools, equipment and the general working practices, which influences actions and provokes errors (Josephson and Hammarlund 1999, p.682; Love *et al.* 2007, p.13). Errors may also arise from error-provoking conditions (or traps) within the research itself.

Although beyond the scope of this paper, it is worthwhile emphasising that any errors committed during research will result in rework if they are identified. Love, Edwards and Smith (2006, p.247) collectively define rework as: "the unnecessary effort of redoing a process or activity that was incorrectly implemented the first time". Rework is only performed when the benefits of performing it is deemed to outweigh the ramifications of not performing it (Manavazhi 2004, p.48). The additional risks of research task interruptions to address rework are that individuals may either forget what it is they were doing or skip performing particular steps in an attempt to compensate for the resultant delay (Love *et al.* 2007, p.12). Undetected errors, of course, remain as a "ticking bomb" for the research and its outcomes.

The following section describes an investigation into data input errors for a questionnaire survey research instrument.

THE RESEARCH INSTRUMENT

A questionnaire was developed as the main instrument to collect data for PhD research that investigates the evolution of workplace architecture as a consequence of technological development in Australia. Assessments were performed to determine whether the questionnaire should either be paper based and mailed to participants or applied online using a web based interface. The online survey offered clear advantages in its distribution, application, data management and overall administration process. The data entry checking process, as explained in this paper, is redundant when using online surveys as the inputted data would automatically be transcribed into the computer.

However, two main factors vetoed the use of online surveys. Firstly, an important aspect of the survey was to assess the level of adoption of technology by companies and map it against demographic variables. The Australian Bureau of Statistics (ABS) reported that in 2002 (latest available report of its kind to date), 99% of companies with one hundred or more employees had Internet connection, but only 65% of companies with less than five employees had this technology. Thus, if the survey was applied wholly online, bias would be introduced as those companies that operate without internet access would not be represented in the online survey. Moreover, even if a company had Internet connection, respondents who are more familiar and comfortable with the Internet will be more likely to respond than those who struggle with or shun this technology. Secondly, the catalogue from which the stratified sample was to be taken included only postal addresses (no e-mails).

A hard-copy of the questionnaire was therefore posted to a stratified sample of 284 participants. After a ten week period, 105 completed questionnaires were received (37% response rate).

The research instrument comprises 36 questions. Almost all of these questions are close-ended, thus respondents were requested to tick the box next to their preferred answer from a list of pre-defined possible answers. Next to each box is a numerical code that refers to a specific answer (e.g. 1=Yes, 2=No, and non-responses are assigned a ‘0’ value). The code corresponding with the selected answer is then input into the computer for subsequent analysis. Although the number of questions per questionnaire is 36, respondents could select more than one answer in some questions. As a result, each questionnaire has a total of 67 entries. Since 105 questionnaires were returned, the total number of data entries is 7,035 (105 x 67).

The likelihood and the consequences of introducing error whilst transcribing just over seven thousand data entries were deemed a high risk to the research. This risk relies on the principle that the research conclusions will be based on the analysis of the data input into the computer, not on the actual data collected. Hence, for these two sets of data to be identical, human reliability (in terms of the propensity to make errors) should be managed.

IDENTIFYING, CONTAINING AND QUANTIFYING INPUT ERRORS

In order to ensure that data inputted into the computer were identical with those on the returned questionnaires, a simple and cost-efficient system was developed. This system utilises both the interpretative capability of humans and the data checking capacity of computers. Figure 1 outlines the developed system implemented not only to contain, but also to quantify the occurrence of input errors whilst transcribing a hard-copy questionnaire into the computer.

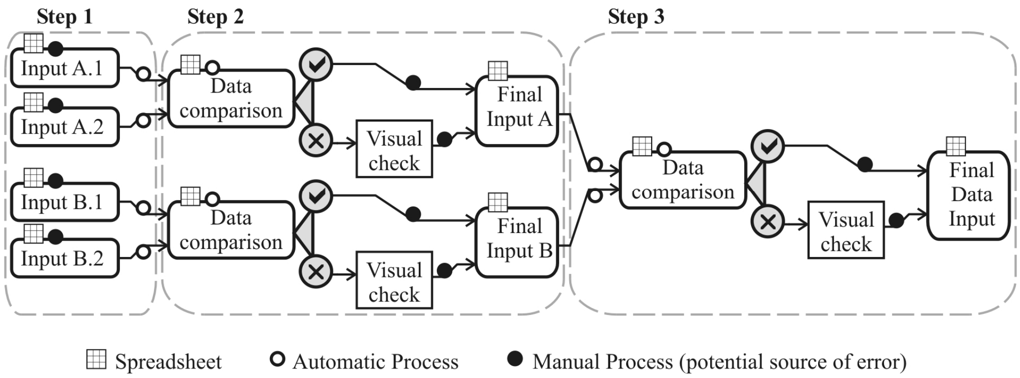

Figure 1: Input process

Although every received hard-copy questionnaire was scanned, the manner in which the answers were registered by respondents made it difficult to automate the reading process (e.g with optical character recognition – OCR – technology). Representative responses that highlight the need of a person, rather than a computer, to decipher the questionnaires are illustrated on Table 1.

Table 1: Representative problems for automated data reading

Type	Example
Ambiguous responses	35 Your gender 1 Male (less) 2 Female
Amended responses	H) General spatial arrangement Poor 1 2 3 4 Good
Illegible responses	5 Approximate percentage of time you spend in the following working modes: (Should add to 100%) 1 % Individual work, and quiet thinking 2 % Face-to-face collaboration 3 20 % Building relationships, and socialising

Step 1: Data Input

Using a spreadsheet, a matrix was created listing all the questions in the first column and all the questionnaire identification numbers (ID) in the first row. The response code (the number next to the tick box) was input in between these two to relate respondents to their answers. Given that this step was performed by two persons (A and B), a procedure was developed and strictly followed in order to reduce errors committed as a consequence of each person adopting different approaches. Questionnaires were input in exactly the same order by Person A and B.

Step 2: Data comparison

A third spreadsheet, 'DataCheck' (refer to figure 1) automatically compares each of the values of 'Input 1' and 'Input 2', which represent both inputs from the same person. If the values are the same (x), it returns a "True" value. If one of the values is different, it returns a "False" value. Table 2, illustrates the system's logic.

Conditional formatting (green-True / red-False) made it easier to identify which entry values, if any, were different. However, the system does not reveal which value is wrong, hence a visual check using the scanned questionnaire was performed to identify whether Input 1 or 2 is correct. Once the input errors are identified, correct values are manually copied and pasted into the 'Final Input' spreadsheet. Being a manual process this is a potential source of input error in itself.

Table 2: Possible data comparison scenarios and results

Input 1	Input 2	Value
X	X	True
X	Any value other than X	False
Any value other than X	X	False
Any value other than X and different from Input 2	Any value other than X and different from Input 1	False
Any value other than X and same as Input 2	Any value other than X and same as Input 1	True*

The last row of Table 2, marked with an asterisk ('*') denotes any embedded flaws within the system. The system will return a 'True' value as long as the two sets of inputs are identical. However, a true value will be misleading when the person performs the exact same error twice. Whilst this is the least probable form of input error to be committed, it creates a degree of uncertainty because a 'True' value in this instance does not always mean a correct value. This is because the system only reveals that the second data input truly matches the first data input, and will not match the correct values provided by the respondents themselves in the returned questionnaires when the same input errors are committed for the same question(s) in both data inputs.

Step 3: Cross data check
The objective of this step is to reduce as far as possible, the probability of two identical errors, as noted above, entering the 'Final Input' spreadsheet. However, it is important to note that whilst this system ensures a reduction in input errors, it cannot ensure that these are eliminated entirely, as there will always be the possibility, however small, that Person A and B will commit the same input error.

This step compares values between 'Final Input A' and 'Final Input B' and automatically identifies differences between the two (refer to Figure 1). As per Step 2 above, conditional formatting is used where the values within the green cells are in this instance deemed correct and those values within the red cells are visually checked against the scanned questionnaires returned.

THE RESULTS

Quantity of input errors

The implemented input process and data comparison system identified a total of 138 input errors incurred whilst transcribing data from the paper-based questionnaires into the computer. Figure 2 illustrates the number of input errors identified during each stage.

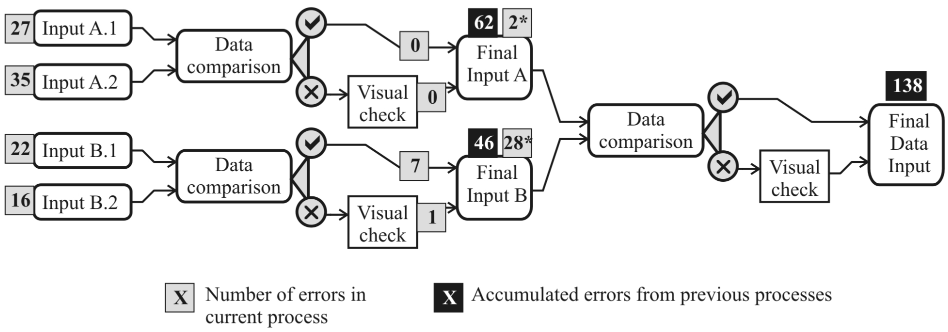

Figure 2: Identified quantity of input errors

A total of 62 identified input errors were accumulated in Input A (Step 1), of which 27 (44%) were committed during Input A.1 and 35 (56%) during Input A.2. In parallel, Input B (Step 1) totalled 38 input errors, of which 22 (58%) were committed during Input B.1 and 16 (42%) during Input B.2. As there was little difference between the percentages of errors committed during Input 1 vs. Input 2, and because the greatest quantity of input errors alternated between Input 1 and 2 for person A and person B, it was inconclusive which of these two sets of inputs was the more reliable. During the copying and pasting of data between spreadsheets Person B committed 8 errors.

The quantity of errors illustrated in Figure 2 with an asterisk (‘*’) alongside them are those input errors identified during Stage 3. As previously discussed, these are errors that were committed twice (once on each Input A/B.1 and Input A/B.2 spreadsheet). To reflect this, the numbers of errors are multiplied by two. Thus, the one input error committed by Person A and the 14 committed by Person B resulted in 2 and 28, respectively. The total number of identified input errors by each person across all stages is 64 for Person A and 74 for Person B. Again, the difference of 10 errors is relatively small when compared with the total amount of data entered.

Analysis of input errors committed per set of received questionnaires reveals that Person A reduced his rate of input errors committed as time passed by, registering 32 errors in the first third of questionnaires (1 to 35), 16 errors in the second third (36 to 71) and 14 errors in the last third (72 to 105). On the other hand, Person B accumulated 14 errors in the first third, 22 in the second and 2 in the last third. Refer to Figure 3.

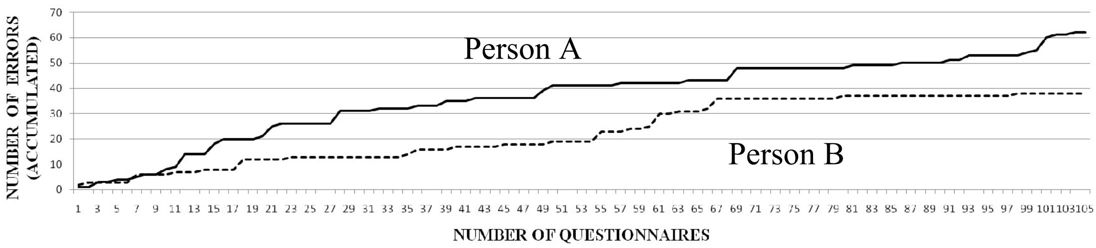

Figure 3: Plot of accumulated input errors committed during Inputs A and B

It is difficult to assert that this represents a “learning curve” effect on data entry reliability, but this may be so, coupled with increasing confidence in personal judgement when encountering problems such as those shown in Table 1.

The total number of input errors committed, 138 for both sets of data entries, is quite small when compared with the overall number of data entries. As previously mentioned, each of the 105 questionnaires had 67 data entries; given that each questionnaire was input twice by Person A and twice by Person B, the total number of entries is 28,140 (105 x 67 x 2 x 2). Thus, these 138 input errors represented only 0.5% of the total data entered.

Still, if no data comparison had taken place and the worst case scenario for Input A (35+1=36) and Input B (22+14=36) were instead hypothetically taken into consideration, the percentage of errors in both cases would have also been 0.5% (of 7,035), but there would have been no way of quantifying these hypothetical errors, let alone containing them. It is noted that for this calculation, errors committed whilst copying and pasting data from one spreadsheet to another were not taken into consideration as they only occurred as part of the data checking process. Table 3 summarises errors by stages.

Table 3: Accumulated input errors committed during Inputs A and B

	Person A	Person B	Total
Input 1	27	22	49
Input 2	35	16	51
Copy and Paste across spreadsheets	-	8	8
Final Input	2	28	30
Total	64	74	138

Impact of input errors

As previously mentioned, the total number of identified errors is 0.5% of the total input data entries. However, for several questions more than one input error was also committed for the same question. It was subsequently discovered that the questions where these input errors occurred had more impact, and were of greater relevance, on the findings revealed from each question than the absolute number of committed input errors identified.

Given that 105 questionnaires were returned, each question had 105 answers related to it. If one answer was inputted incorrectly 0.95% impact of the input error is introduced. Similarly, if two responses within the same question are incorrect, the impact percentage of the input errors increases to 1.90% and so forth. Table 4 illustrates that Person A committed 1 input error per question for 18 questions, 2 input errors per question for 10 questions, 3 input errors per question for 5 questions, and 4 and 5 input errors per question for 1 question each. In comparison, Person B had committed 1 input error per question for 27 questions, 2 input errors per question for 4 questions, and 3 input errors per question for 1 question. The column furthest right in Table 4 illustrates the impact percentage of the input error(s) introduced to the same question.

Table 4: Quantity of input errors committed per question

Quantity of input errors in the SAME question	Person A	Person B	Impact % of input error
1	18	27	0.95%
2	10	4	1.90%
3	5	1	2.86%
4	1	-	3.81%
5	1	-	4.76%
Total	18x1+10x2+5x3 +1x4+1x5=62	27x1+4x2+1 x3=38	

As errors familiar in nature continue to occur within any system, the tendency is for “recurrent error traps” to materialise (Love *et al.* 2007, p.13). The impact of this is identified by Reason (2002, p.42) who likens error traps to a simple photocopier machine and demonstrates that it is indeed possible to establish these error traps (Reason 2002, p.43). Under this context, one question in the questionnaire was identified as an error trap as it registered 6 errors overall, 3 by Person A and 3 by Person B. Therefore, the questions where most of the input errors were committed represent error traps as they contain input error-provoking conditions. Additionally, the impact of input errors committed per question is further magnified when a correlation between two variables is performed.

Time

In order to assess whether a relationship exists between the time taken to input data and the quantity of input errors committed (i.e. do faster inputs produce more errors?), the time taken by each person to input data from each questionnaire received was monitored and analysed. Table 5 displays the data input times by Persons A and B for Inputs 1 and 2 and for each lot of questionnaires returned (1-35, 36-71, 72-105).

Table 5: Inputting time (HH:MM:SS)

	Input A1	Input A2	Input B1	Input B2
1st Third (1-35)	01:53:00	01:29:00	01:18:00	01:00:00
2nd Third (36-71)	01:28:00	01:12:00	01:06:00	00:55:00
3rd Third (72-105)	01:11:00	01:02:00	00:45:00	00:45:00
Sub-total (by Input)	04:32:00	03:43:00	03:09:00	02:40:00
Sub-total (by Person)	08:15:00		05:49:00	
Total	14:04:00			

From the above table it can be seen that the overall input time is just over 14 hours, with a difference of 2 hours and 26 minutes between Input A (8:15hrs) and Input B (5:49hrs). Table 5 also illustrates a decrease in keying time between the first and the second lot of questionnaires and a decrease again between the second and third. This apparent trend for Inputs 1 and 2 of Persons A and B, suggests that they both became more dexterous in keying data as time progressed.

Another consistent trend is that the keying time of Input 2 was in all cases faster than Input 1. This could indicate that: a) more 'care' was taken whilst keying Input 1, or b) given that the second input immediately followed the first input, some of the responses may have been recalled and input faster, or c) a combination of both of these. Table 5 also shows that Person B achieved the greatest level of dexterity on the last third of questionnaires (45 minutes), averaging one second per input of single data (reading and typing).

The times displayed in Table 5 are those taken to actually input the data. However, the complete data entry process was performed over several days. Referring to Figure 4, Person A completed his input in 6 days. Day breaks, denoted with thick vertical lines, were taken after the input of questions 21, 25, 50, 85 and 89. Three breaks with a minimum duration of two hours, denoted with thin lines, were taken by Person A after the input of questions 18, 40 and 104. Referring to Figure 4, Person B completed his input in 3 days, with day breaks taken each time after the input of questions 22 and 44. Only one two hour break was taken by Person B after the input of question 38.

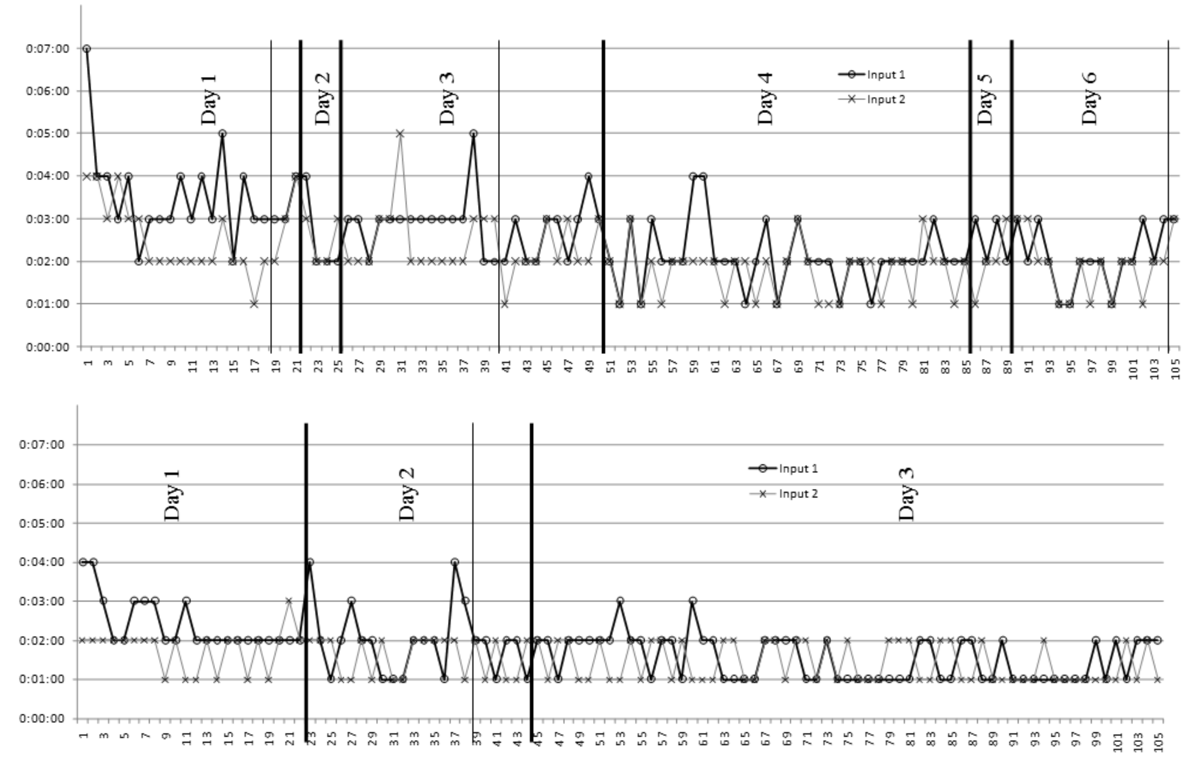

Figure 4: Plot of times for Inputs 1 and 2 for Person A (above) and Person B (below)

After comparing Figure 4 with the accumulated error plot (Figure 3), it can be observed that Person A did not achieve the same level of proficiency as Person B. The former not only took longer and showed a more chaotic inputting rhythm, but also committed more mistakes.

CONCLUSIONS

Studies that rely on paper-based surveys to collect data have the embedded risk of introducing error to the research conclusions due to deficiencies in human reliability at the data entry stage. In order to reduce this risk, appropriate measures should be adopted to ensure that the data registered by respondents on the survey instrument is identical to that inputted within the system tool (i.e. software) used to analyse such data.

The consequences of a relatively small impact percentage of the input error(s) committed (0.5% in the experiment) in transcribing data can affect the overall research findings (~5.0%). Understanding where to direct the bulk of error containment focus will therefore benefit the research deliverables. Addressing the error traps within the research instrument itself will lead to improvement in the reliability and credibility of the research outcomes. Questions where the majority of all input errors are committed represent error traps within the research as they contain conditions that provoke input errors.

Although it is suggested that online surveys eliminate altogether this type of error as well as the time required for transcribing data, this may not be a feasible means of data collection for every form of research. If insufficient care is taken to adopt an appropriate means of data collection, this can potentially introduce other sources of input error (e.g. research bias) which may be more difficult to manage than those input errors managed within this research.

ACKNOWLEDGMENTS

This paper is part of an ongoing PhD investigation on 'The Evolution of Workplace Architecture as a consequence of Technology development in Australia' being undertaken by the first author and sponsored by the Australian Cooperative Research Centre for Construction Innovation (CRC-CI). The authors would like to thank the School of Property, Construction and Project Management at RMIT University especially Mr. Ehsan Gharaie, A/Prof. Helen Lingard and Prof. Ron Wakefield for their time, comments and contribution to this paper.

REFERENCES

Busby, J.S. (2001) Error and distributed cognition in design. *Design Studies*, **22**(3), 233-254.

Josephson, P.-E. and Hammarlund, Y. (1999) The causes and costs of defects in construction: A study of seven building projects. *Automation in Construction*, **8**(6), 681-687.

Kaminetzky, D. (1991) *Design and Construction Failures: Lessons from Forensic Investigations*. McGraw-Hill, New York.

Love, P.E.D., Edwards, D.J., Irani, Z. and Walker, D.H.T. (2007) Project pathogens: the anatomy of omission errors. *IEEE Transactions on Engineering Management*, (In press)

Love, P.E.D., Edwards, D.J. and Smith, J. (2006) Contract documentation and the incidence of rework in projects. *Architectural Engineering and Design Management*, **1**, 247-259.

Reason, J.T. (2002) Combating omission errors through task analysis and good reminders. *Quality and Safety in Health Care*, **11**(1), 40-44.

Reason, J.T. and Hobbs, A. (2003) *Managing Maintenance Error: A Practical Guide"*, Ashgate Publishing Company.

Sunyoto, A. and Minato, T. (2003) Representing causal mechanism of defective designs: a system approach considering human errors. *Construction Management and Economics*, **21**(3), 297-305.