Adding Context to Automated Text Input Error Analysis with Reference to Understanding How Children Make Typing Errors

by

Akiyo Kano

School of Computing, Engineering and Physical Sciences

University of Central Lancashire

Preston, UK

A thesis submitted in partial fulfilment for the requirements of the degree of Doctor of Philosophy at the University of Central Lancashire

November 2011

Student Declaration

Concurrent registration for two or more academic awards

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution.

Material submitted for another award

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work.

Collaboration

Where a candidate's research programme is part of a collaborative project, the thesis must indicate in addition clearly the candidate's individual contribution and the extent of the collaboration. Please state: None.

Copyright

Signature of Candidate	___________________________
Type of Award	Doctor of Philosophy
School	Computing, Engineering and Physical Sciences

Adding Context to Automated Text Input Error Analysis with Reference to Understanding How Children Make Typing Errors

Akiyo Kano

School of Computing, Engineering & Physical Sciences

University of Central Lancashire

Preston

akano@uclan.ac.uk

ABSTRACT

Despite the enormous body of literature studying the typing errors of adults, children's typing errors remain an understudied area. It is well known in the field of Child-Computer Interaction that children are not 'little adults'. This means findings regarding how adults make typing mistakes cannot simply be transferred into how children make typing errors, without first understanding the differences.

To understand how children differ from adults in the way they make typing mistakes, typing data were gathered from both children and adults. It was important that the data collected from the contrasting participant groups were comparable. Various methods of collecting typing data from adults were reviewed for suitability with children. Several issues were identified that could create a bias towards the adults. To resolve these issues, new tools and methods were designed, such as a new phrase set, a new data collector and new computer experience questionnaires.

Additionally, there was a lack of an analysis method of typing data suitable for use with both children and adults. A new categorisation method was defined based on typing errors made by both children and adults. This categorisation method was then adapted into a Java program, which dramatically reduced the time required to carry out typing categorisation.

Finally, in a large study, typing data collected from 231 primary school children, aged between 7 and 10 years, and 229 undergraduate computing students were analysed. Grouping the typing errors according to the context in which they occurred allowed for a much more detailed analysis than was possible with error rates. The analysis showed children have a set of errors they made frequently that adults rarely made. These errors that are specific to children suggest that differences exist between the ways the two groups make typing errors. This finding means that children's typing errors should be studied in their own right.

ACKNOWLEDGEMENT

Like all theses, this thesis was the product of contributions by many people. I am grateful that these people are in my life and were all kind enough to help me.

Without my mother, I would have never found myself in England, let alone do a PhD. She has always believed in me and supported me in everything. She has taught me that I could do whatever I wanted regardless of any disabilities. She is a true inspiration to me, and I will forever be thankful that I am her daughter.

My Director of Studies, Dr Janet C Read deserves a special mention for being the most understanding supervisor one could ask for. Her ability to know when a student requires a break and when they need a push was the key in the completion of this thesis. Her encouragement to write for and attend conferences drove this research to a much higher level.

Additionally, I was lucky enough to have received advice from Prof. Alan Dix of Lancaster University, who is a fountain of ideas and inspirations. Prof. Scott MacKenzie of York University, taught me that I had 'too many variables' and ruthlessly advised me on how to write academic papers. I would also like to thank Dr. Jacob Wobbrock of University of Washington and Dr. Petros Roussos of University of Athens for kindly sharing their work with me.

I am very fortunate to be part of a thriving research group. I am indebted to my friend and colleague Nicky Danino for being my advocate, keeping me sane, and in good humour throughout the process. Gavin Sim for teaching me to look at things from different perspectives. Matthew Horton for assisting me in many of the data collection sessions. Lorna McKnight and James E Thompson for their friendships and giving me the support I needed to finish the thesis.

Finally, this thesis would not been possible without the participation of nearly 600 children and students. I thank each participant, their teachers and their families for allowing the studies to take place.

Dedication

This thesis is dedicated to my mother, and for all those that keep on achieving in the face of disabilities.

Table of Content

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

1 INTRODUCTION

At the start of the 20th Century, when the field of typing research began, researchers were uninterested in understanding how children typed. This was because children rarely used the typewriter. In those times, typing (or typewriting) was a skill taught to (mostly female) adults as a pathway to gaining employment as typists and secretaries.

However, as the typewriter, and then the personal computer, became more widely used, children began to type. In particular, the introduction of computers in schools in the 1950s (O'Shea and Shelf, 1983) and introduction of microprocessors in the mid 1970s, making computers more affordable, ensured every school child gained experience on the keyboard. In the 2000s, children who were born into a world of ubiquitous digital devices acquire typing skills with no formal training. Computers (and thus typing) have become a core part of the British educational system, which states children must be taught ICT from the first year of their primary school education (DoES, 1989).

Vast bodies of literature exist on analysis of typing errors made by adult typists (Chapter 3). These studies range from gathering the complete spectrum of typing errors, to those focused on a specific few in an attempt to understand the psychological causes of the errors. With research spanning over a century, it is fair to say that typing errors made by adults is a well-studied area. However, children's typing errors remain an understudied area.

Many studies on the benefit of using computers in education (Wood and Freeman, 1932; Rowe, 1959; Krevolin, 1965; Roussos, 1992) argued that the use of computers improves children's reading, writing and understanding of subjects. However, these studies do not look at the difficulties children experience when typing. There is little work on understanding the typing errors made by children. What typing errors do they most frequently make? Do children make the same typing errors as those made by adults? Are there any errors specific to young children? This thesis was motivated by these unanswered questions.

1.1.1 Structure

To introduce the main themes of this thesis, Section 1.2 provides a brief background to typing error analysis, and highlights the lack of studies investigating children's typing errors. Section 1.3 discusses the main issues surrounding comparing children and adults typing, which were addressed in this thesis. Section 1.4 formally outlines the objectives of this thesis and the chapter concludes in Section 1.5 with a summary of the chapter and a brief description of the structure of the remaining thesis.

1.2 TYPING ERROR ANALYSIS

Typing error analysis refers to the study of typing errors made whilst using the keyboard to produce a body of text. Most frequently in such studies, participants are shown a body of text and are asked to type the text using a typewriter (computer keyboard in the later years). The text presented to the participant and what was typed are compared to see if any typing mistakes were made.

Typing error analysis has been popular in two fields - psychology and HCI. Psychologists first started to investigate typing errors in the 1910s, when typewriters became widely used. Since time was money and typing errors cost time to correct, they were interested in minimising these errors. They suggested that to reduce errors, one must study the typing errors themselves to understand the causes (Book, 1925). The earlier studies were largely descriptive, such as counting how many times each letter was substituted for another (Lessenberry, 1928). This soon developed into researchers defining their own set of typing errors (Chapter 3 will discuss these in detail), which were applied to real typing data collected from professional typists. Researchers made guesses as to their cause, and provided workbooks for typists designed to reduce particular typing errors (Opfer, 1932; Schoenleber, 1932).

Up to this stage, comparisons were made only between text that the participant was asked to copy type (presented text - PT) and the final typed text that was produce (transcribed text - TT). However, in the 1960s, a new surge of interest arose from the ability to record timings of each key press. This meant researchers were able to study in detail the input stream (IS), containing every key pressed by the participant. Spearheaded by Donald Norman and David

Rumelhart at the LNR Research Group, San Diego, timing between key presses became the central focus in the modelling of cognitive functions involved in typing. Many of these works are summarised in (Salthouse, 1986).

A third surge in interest occurred with the rising increase in mobile computing devices since the 1990s (MacKenzie and Soukoreff, 2002b). With the explosion of new text input methods designed for these devices, it became important to be able to evaluate one method against another (MacKenzie, 2007). MacKenzie and Soukoreff (Soukoreff and MacKenzie, 2001; MacKenzie and Soukoreff, 2002a; Soukoreff and MacKenzie, 2003; Soukoreff, 2010) have led the development of various measures of error rates - the ratio of typing errors to the amount of typing done. Algorithms were also developed (Soukoreff and MacKenzie, 2001; Wobbrock and Myers, 2006) to enable computation of these error rates with ease. The convenience of automated calculation of error types has made them increasingly popular in HCI. However, Soukoreff (2010) points out that these error rates only tell whether one method or participant is better than another, not *how* or *where* the two differ.

In contrast to the well-studied area of adult typing, little work exists in studying the typing errors of children. Roussos (1992) carried out several typing studies with children for his thesis. In these studies, he counted the number of errors created by the participants as an evaluation method, but did not define how he counted them.

Read et al. (2001) studied short phrases copied by children aged between 6 and 10 years. The children entered text using four different text input methods (mouse, keyboard, speech recognition and handwriting recognition). To evaluate these methods, they defined six error types and classified the typing errors accordingly. They found that children displayed difficulties using the chosen text input methods, such as Execution Error (e.g. the child pressed the key for too long). They also found that children demonstrated 'exaggerated versions' of errors commonly made by adults. They pinpointed sources of error to be at the user, the hardware and the software.

In another study, Read and Horton (2006) found that teenagers carrying out a similar phrase-copying task were more prone to errors in controlling their hand movements, such as pressing the key next to the one they had intended on pressing. Although comparisons between children and teenagers were not made, these two studies by Read show there is a difference in the range of errors the two age groups make. However, the two studies had a small number of participants (12 and 18 respectively).

Additionally, the current literature lacks direct comparisons between children and older participant groups using exactly the same task. It was important to establish whether children made the same range of typing errors to adults or not, since if they did not, the findings from adult studies could not be directly applied to children.

The contribution of this thesis in terms of studying typing is the detailed investigation of the typing errors made by children, and how this differs from typing errors made by adults. This comparison is made in Chapter 10.

1.3 Comparing Children and Adults

For the findings of a comparison study between children's and adults' typing to be valid, it was crucial to ensure both the data collection and analysis methods produced comparable data between the two participant groups.

1.3.1 Seeking a Comparable Data Collection Method

Since children's typing errors have not been compared directly with adults' typing before, it was unknown as to whether or not the methods used to collect typing from adults were appropriate for use with children.

The most frequently used method in gathering typing from adults is to show them a body of text (e.g. magazine articles, business documentation, a set of short phrases) on paper for them to copy. The questions here for children are what text to show them, and how to show the text. Firstly, in order for the text to not cause any bias towards the adults, it was important to ensure the language used and the length of the text is suitable for children. Secondly, it was unknown as to whether or not children would have any difficulties in copying text presented on paper. MacKenzie and Soukoreff (2002b) argued that paper-to-screen copying has a higher Focus of Attention (FOA) than screen-to-screen copying where the presented text is displayed on the same screen as the transcribed text. It was unknown as to whether or not this increase in FOA would cause difficulties in children.

Chapter 2 explores the areas of the data collection method that required further investigation before one could be certain it was suitable for children. Chapter 6 studies difficulties children experience in paper-to-screen copying tasks, which supports the use of the screen-to-screen mode instead. Chapter 8 defines a new phrase set designed to be suitable for use with children aged six years and above.

1.3.2 Seeking a Comparable Data Analysis Method

The selection of the right method to categorise the typing errors was also important. Firstly, to capture all the typing errors, the chosen categorisation method must not exclude any error types. Secondly, to gain an accurate picture of the typing errors made, the method must not break down larger errors into many smaller typing errors. An example of this is to ensure that an omitted word is captured as one omitted word rather than several omitted letters.

Thirdly, since the cause of children's typing errors has not been studied in detail, it was decided that the categorisation method could not make any assumptions as to the causes. Some categorisation methods included error types such as 'error of distraction' (Clem, 1929) and 'deviation from copy' (Book, 1925), which forced assumptions as to the cause of the error.

Finally, the method must not assume that children have formal typing training. In the early half of the 20th century, when most typists were trained to use the touch-typing method (Clem, 1929), they were taught to use particular fingers for particular banks of keys on the keyboard. In studying these typists, it was fair to assume that one could define errors by the use of wrong fingers. Examples of such error types are 'adjacent letters by using wrong finger' (Opfer, 1932) and 'homologous error' (Grudin, 1983a) where the correct finger on the wrong hand was used to type (resulting in a substitution). However, children nowadays are no longer trained in touch-typing. This results in most typists using whatever finger is most convenient to type the letter. It was therefore not appropriate for the method to assume that participants have had formal touch-typing training.

Chapter 3 reviews many of the categorisation methods found in literature. Each was evaluated for suitability for use in this research, but none satisfied all four conditions. Chapter 8 explores the cost of accuracy to the typing data by use of the wrong categorisation methods before defining a categorisation method that did fulfil all the requirements. This new categorisation method is then used in Chapter 10 for detecting and categorising typing errors by children and adults.

The contributions of this thesis in terms of comparing adults and children are the data collection and analysis methods that produce comparable data between children and adults, and defining new error types specific to children's typing.

1.4 THE THESIS

The primary research question of this thesis was '*are there notable differences between typing errors made by children and adults*'. Several objectives helped achieve this primary aim. These were:

- Establishing a data collection method for typing errors that causes the least amount of bias between children and adults.
- Establishing a typing error categorisation method that encompasses the whole range of typing errors made by children, without making any assumptions as to their cause.
- Automating the data collection and analysis process as much as possible to ensure an efficient, consistent and valid study.
- Find any typing error behaviours that are specific to children.

The thesis of this work is: *'there is a set of typing error behaviours that are specific to children in phrase-copying typing'*. Although children and adults share some error types, there are error types that are almost exclusively made by children. This difference suggests that theories on how adults make typing mistakes cannot be applied directly to children without first testing the theories with real children's typing data.

1.4.1 Structure

An investigation of the methods used to collect and analyse typing errors revealed a plethora of methods. Chapter 2 reviews the methods used in other studies to collect typing data. These methods were designed for use with adult participants. Using these adult-oriented methods with both adults and children would have created a bias towards the adults, invalidating the study. A new data collection method was required that could produce as comparable data as possible between children and adults.

A new phrase set, designed to be suitable for participants above the age of six years old is introduced and evaluated in Chapter 5. The method of presenting these phrases to the participants on paper is evaluated in Chapter 6, and rejected in favour of presenting them on screen. Two questionnaires (one for adults and one for children) designed to gather participants' previous computer experiences are evaluated in Chapter 7.

A second literature review (Chapter 3) surveyed the analysis methods used in typing error studies. It revealed a similar lack of consideration of child typists in these categorisation methods. 22 categorisation methods were evaluated but none fulfilled all the criteria for categorising typing errors made by children. The existing methods were either thorough but made causal assumptions, or only defined a select few error types. Furthermore, only one of these methods had ever been applied to young children's typing. Therefore, it was necessary to design a new categorisation method based on real typing errors made by children (Chapter 8).

Using the categorisation method manually was hugely time consuming and was prone to inconsistencies and ambiguities. To address these issues, a program was developed to carry out the detection and categorisation (Chapter 9). This program is an extension of existing automated analysers. The existing analysers were only able to carry out letter-level analysis and lacked the contextual information such as whether an error was part of a consecutive group of errors or not. The new analyser is able to use contextual information to provide much more meaningful categorisation. It also carries out some basic disambiguation tasks to reduce ambiguities in a consistent manner.

The study described in Chapter 10 uses all the tools designed and evaluated in the thesis to compare the typing errors made by children and adults. It found that children and adults do make a different range of error types, and identified typing error behaviours that were unique to children.

1.5 CONCLUSIONS

This chapter has outlined the motivation for this thesis and the research objectives it set out to achieve. The research was motivated by the lack of work in investigating typing errors made by children and comparing them to those made by adult typists. This thesis aimed to establish new methods that allow comparable typing data to be collected and analysed from children and adults to discover whether or not there are any major differences between the two groups of typists.

The remainder of this thesis comprises of the first three chapters describing the literature review carried out and considerations made to the research methodology used in this research. The three chapters following this design and evaluate a new data collection method that is more valid for use with both children and adults than previous methods. This is followed by a chapter on creating a categorisation method based on typing errors made by children as well as those made by adults, already observed in literature. The process of automating the detection and categorisation of typing errors is then discussed, allowing for a much faster, more accurate and consistent analysis of typing errors. Finally, a large-scale study uses these tools to answer the primary research question.

2 ISSUES RAISED IN DESIGNING TYPING ANALYSIS EXPERIMENTS FOR CHILDREN

2.1 INTRODUCTION

Typing is a complex task. Cooper (1983b) defined typing in terms of five stages - character recognition (reading the words to type), storage buffer (memorising the words), motor program (instructing the finger to move), the keystroke (pressing the key) and the sensory feedback (checking that the execution of typing was done correctly). Since it is difficult to study all these aspects in a single study, researchers have concentrated their studies on small areas of typing at a time. This focused approach to studying typing created many different experimental methods, with no single standard method. This meant that a text input researcher must carefully select a method most suited to their aim through the wide range of methods available.

Comparing the typing of adults and children has the added complication in that children have less developed cognitive, linguistic and motor abilities. This means extra attention must be paid in designing the task so it is not unfairly taxing on the children (Markopoulos et al., 2008).

This chapter reviews the range of experimental design options found in literature to deal with issues surrounding the typing task. Discussion is made on each issue as to what is most suitable for addressing the research question of this thesis of detecting notable differences between the way children and adults make typing mistakes. Some issues can be resolved using existing methods, but in other issues, it was unclear as to which method was most suitable for children.

2.1.1 Objectives

The purpose of this chapter is to introduce the typical typing tasks used in text input analysis and highlight some issues in using these tasks with children. It focuses on the decisions made by other researchers who studied adult typing, discussing the advantages and disadvantages of each choice. It then questions which method is most suitable for use with children, and highlights areas that required further investigation.

2.1.2 Scope

This literature review focuses on the experimental designs used in studying the typing errors, discussing the different options a researcher could take in such studies, and the effects of the decisions in the result. Since the field of text input is vast, several constraints were placed on this literature review.

Many historical reviews of the field of text input analysis already exist, so this review will not attempt to provide duplication. For a thorough and wide-ranging historical review of the typewriter and many other forms of text input methods ranging from the first typewriter (1873) to the 1970s, the reader is directed to Yamada (1980). Silfverberg (2007) provides a similar review but focused more on modern text input methods. Other historical reviews of text input methods include (Russon and Wanous, 1973; Cooper, 1983a).

As this thesis focuses solely on the full-size keyboard, this review is restricted to studies on full-size keyboards. Although studies on reduced keyboards and other forms of text inputs are mentioned here, they were done so for the decisions made in the experimental design rather than for the particular input device being studied. For a survey of mobile text input evaluation methods, the reader is directed to (MacKenzie and Soukoreff, 2002b).

This thesis also focuses on the use of the QWERTY keyboard layout only. For a historical review of the QWERTY keyboard and other keyboard layouts, please see Noyes (1983). Many other studies have also compared the QWERTY keyboard to other keyboard layouts (Davis, 1935; Janelle, 1974; Hopkins, 1989; Joyce and Moxley, 1989; Buzing, 2003).

2.1.3 Structure

Section 2.2 outlines the motivations in studying how people type. The chapter then narrows its focus down to comparing how adults and children differ in their typing. Section 2.3 discusses experimental design options in literature and evaluates their suitability for use with children. The chapter concludes in Section 2.4 with an outline of the experimental design decisions already made, and what issues remain.

2.2 MOTIVATION FOR STUDYING TYPING PERFORMANCE

Researchers have been carrying out text input studies since the widespread use of the typewriter in 1870s (Yamada, 1980) to the current day. In this thesis, text input analysis refer to studies investigating the typing speed and/or typing errors of text typed by participants. In these studies, the participants are shown a text to copy and asked to produce typing through a text input method(s). The text input method could be anything from a full-size QWERTY keyboard, numerical keypads such as those found on a mobile phone, or stylus input, to soft keyboards displayed on a screen.

There are four main reasons for studying typing (author's list):

1. To understand the causes of typing errors
2. To understand psychological processes in typing
3. To evaluate a new text input method
4. To select a suitable text input method from a range of methods

2.2.1 Understanding the Cause of Typing Errors

At the start of the widespread use of typewriters, the emphasis was on increasing the typing speed. However, the focus soon shifted to improving the accuracy. Error correction on a typewriter was costly in time and appearance, so it was desirable to minimise typing errors. Early researchers thought that to reduce typing 'demons', these demons must first be understood (Ford, 1928). These studies attempted to gain an overview of typing errors by categorising every error made in a typical typewriting task (Book, 1925; Ford, 1928; Clem, 1929). Such works include tabulating frequency of what letter was substituted for what letter (Lessenberry, 1928), defining and counting all typing errors found (Book, 1925; Opfer, 1932; Schoenleber, 1932) and studying frequency of errors in the most commonly used words (Ford, 1928). Some of them made guesses as to the causes of these typing errors. These works were often accompanied by workbooks full of instructions and exercises on how to correct each type of error (Book, 1925; Rowe, 1931; Opfer, 1932). However, it was not possible to understand the real cause of the typing errors until the psychological processes behind typing could be understood.

2.2.2 Understanding the Psychological Process of Typing

Between the 1960s and 1980s, psychologists made significant advances in constructing models of the psychological processes that took place during transcription typing (copying a text). The tasks of reading a text and typing the words were deconstructed into smaller components and studied carefully.

These studies did not use the typical typewriting tasks but used tasks designed to answer specific questions about the psychological processes. For example, to understand the importance of visual and auditory feedback in touch-typing, Diehl and Seibel (1962) covered the keyboard and masked the sound of the keys being pressed. Researchers painstakingly manually combed videos and key logs to find measurable details such as the time taken between key presses, motion of the fingers and reaction times. These data surrounding the 'during' action of typing allowed the researchers to draw models of typing in much more detail than the transcribed text alone ever could.

For example, several studies found that successive keystrokes by alternate hands are faster than those made with fingers from the same hand (Fox and Stansfield, 1964; Gentner, 1982; Norman and Rumelhard, 1983; Salthouse, 1984). They argued that this was because when the successive keystrokes are on the same hand, there is little chance for preparing the second keystroke in advance, whereas when on different hands, the second hand could prepare to press the second key before the pressing of the first key was completed. This observation was used to argue against a serial order model that stated letters were processed in the mind one by one (Lashley, 1951).

In all of these studies, timing of key presses immediately before, during and after the error were seen as important indicators of difficulties. However, the timing of key presses were measured by electric timers only accurate to the nearest second or two (Diehl and Seibel, 1962).

The arrival of the electric typewriter and later the computer keyboard made error correction less costly. There was no longer such an urgent need to reduce typing errors. However, timing of key presses could now be measured in fractions of seconds, providing researchers with more details than ever before. This resulted in a second surge of research investigating the time taken to type

(Isokoski and Raisamo, 2000), intervals between key presses (Hiraga et al., 1980; Inhoff, 1991) and reaction times (Hayes, 1978) in much more detail.

Differences among the participants, particularly that of typing skill, were also studied (Fendrick, 1937; Provins and Glencross, 1968; Grudin, 1983a; Salthouse, 1984). The aim here was to see how varying typing skills affected typing performance. Typically, participant groups differing in typing skills were asked to carry out a task (or set of tasks), using the same input device. The typing performances between the groups were compared to draw conclusions.

2.2.3 Evaluating New Text Input Methods

In more recent years there has been an explosion in computer technology. What was once dominated by the personal computer and the full-size keyboards has now seen the introduction of mobile, surface and tablet devices (MacKenzie and Soukoreff, 2002b). The first questions often asked of these new input methods are 'how fast is it?' and 'how accurate is it?' (Wobbrock, 2007). Therefore, the most dominant reason for carrying out a text input study nowadays is to evaluate a new text input method (Davis, 1935; MacKenzie, 2002b; Oniszczak and MacKenzie, 2004). MacKenzie (2007) states that these comparison studies should ask the following questions:

- Will they support fast and accurate text entry?
- Will they do so with modest practice or do they require prolonged practice?
- What are the expected entry speeds or error rates?
- Will users like the techniques or will they find them frustrating?
- How do they compare to alternative text entry techniques?
- Are there aspects of the designs that could be modified in some way to improve performance?

In studies evaluating a particular text input method, a second text input method is chosen to compare with. A typical task (or a range of tasks) is selected for the participants to perform on both input methods. A metric is then used to see how the two input methods compare with each other.

2.2.4 Selecting a Suitable Text Input Method

Another situation in which text input methods may be studied is in selecting a suitable text input method for a device (MacKenzie et al., 1994; Lewis, 1999; Butts and Cockburn, 2001; Zhai et al., 2005). A range of text input methods may be evaluated through various tasks typical to the device. Here, the focus is on selecting the one text input method most suited to the typical user doing the typical tasks for the device.

Although a wealth of research exists in comparing text input methods for adults, only a handful have been carried out with children. Read et al. (2001) compared the usability of four text input methods (mouse, keyboard, speech and handwriting recognition) for children aged between 6 and 10 years. They found that children seem to have a higher tolerance to errors than adults did. Roussos (1992) carried out several studies comparing full-size QWERTY keyboard layout with alphabetic layout for children aged between 6 and 8 years old. He found that the QWERTY layout put an additional load on the user's memory.

2.2.5 Motivation of this Thesis

In this thesis, the aim is closest to the aim of understanding the psychological processes of typing. However, the focus of the study here is in the detection of the differences that may rise in the assumed psychological and cognitive differences between young children and adults. The pursuit of the actual cause of these observable differences is outside the scope of this thesis.

Although child development theories such as Piaget (1896-1980) and Vygotsky (1896-1934) approach the process of cognitive development of a child differently, they all agree that a healthy child's thinking, language, memory, problem-solving abilities and reasoning increase with age. This implies that young children are cognitively different from fully developed adults.

Compare this with the five stages of typing defined by Cooper (1983b):

1. Character recognition (reading the words to type)
2. Storage buffer (memorising the words)
3. Motor program (instructing the finger to move)
4. The keystroke (pressing the key)

5. The sensory feedback (checking that the execution of typing was done correctly)

Typing is a task requiring many of the cognitive functions. Therefore, it is reasonable to assume that, with the cognitive differences between children and adults, there will be notable differences in the way the two groups make typing errors.

There is an interesting debate as to who constitutes as a child. The United Nations Conventions on the Rights of the Child (UNCRC) defines children as anyone under the age of 18 years. Legal upper ages of 'minors' around the world vary from 16 to 21 years. However, age of criminal responsibility is 10 in England and Northern Ireland, and 12 in Scotland.

This thesis is concerned with the differences between young children and adults, so for the purpose of the thesis children were defined to be those attending primary school (3 to 11 years old). In the pilot study described in Chapter 6, it was decided that children under the age of seven were unsuitable for participation so the age of participants was narrowed down to between 7 and 11 years of age. Primary school children were chosen as the 'children' participants, simply because there is a more contrasted difference in terms of cognitive ability and computer skills between primary school children and adults, than between secondary school children and adults.

2.3 THE TYPING TASK

Once the aim of the experiment is defined, the experimental design has to be carefully considered. In designing an experiment, the independent variable is chosen (in the case of this thesis, whether the participant is a child or an adult). All efforts must be made to ensure that the other variables remain constant.

Decisions regarding experimental design must be based on validity - whether or not choosing method A over method B makes the finds more valid. In typing studies, careful considerations are required in choosing the right method so that 1) the effects observed on the dependent variables originate solely from the changes made in the independent variables (internal validity) and 2) the findings are generalisable (external validity).

In any experiment, the apparatus (device), the participants, the task and the analysis method require careful consideration (Figure 1). All four factors must be taken into account when designing a typing experiment valid for both adults and children.

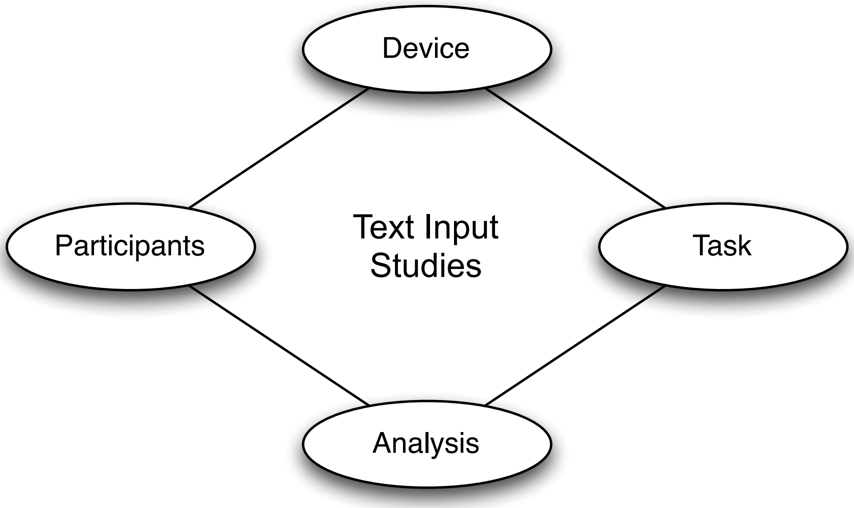


Figure 1: Four Factors of Text Input Studies

When one of the four aspects is chosen for manipulation, the other three aspects must stay constant. Since the thesis is focused on comparing different participants, the device, the task and the analysis method's effect on the outcome must be kept at a minimum. Keeping the device constant is relatively easy. One must simply ensure that all participants type on the same design of keyboard - in this case, a full-sized QWERTY keyboard with an UK Windows layout (see Appendix 1 for the layout), with white writing on black keys.

However, keeping the typing task and the analysis method's effects on the results to a minimum is much harder. The remainder of this section will outline the experimental design questions surrounding the task (the effect of analysis method will be examined in Chapter 8). The assumption made in the following section is that the researcher has already made the decision to carry out a typing study using a full-size QWERTY keyboard to compare the performance of participants. For a detailed guide on how to design a comparison study of different text input methods, readers are directed to MacKenzie (2007).

2.3.1 Create Text From Scratch or Copy Text

Researchers who gather typing data have made a clear distinction between text-creation and text copy (MacKenzie and Soukoreff, 2002b). Text-creation refers to letting the participant type 'whatever comes to mind'. In this method, the participant would be asked to literally type whatever came to their mind, or perhaps be given a topic to write about. This method has the highest external validity since it is close to the typical usage of text input with text natural to the participant. However, this method is reliant upon the participant being able to type a long piece of text under test conditions. Although this is not of serious concern in adults, young children may have difficulties in producing text under pressure. Additionally, when a typing error is found, it is impossible to know what the participants really intended to type (MacKenzie and Soukoreff, 2002b). Knowing what the participant intended to type is crucial in categorising the typing error. Furthermore, since the participant only uses words they are familiar with, there is no control in the distribution of letters and words entered (MacKenzie and Soukoreff, 2003). External validity dictates that the letter and word distributions should be as close to the language as possible. A person may avoid using certain words, skewing this distribution.

Text-copy refers to showing the participant a body of text or short phrases to copy. Although this method lowers the external validity somewhat by restricting what words are used, it overcomes the disadvantages of text-creation. Firstly, there is no time and cognitive cost in trying to create a new text. Secondly, it is easy to estimate what the participant intended to type. Finally, control of the distribution of letters and words is much easier, and can be tested for their closeness to the natural language in advance (MacKenzie and Soukoreff, 2003).

Text-copy offers a good middle ground between trying to capture natural, creative typing and the control required to compute useful data from the typing. This makes text-copy ideal for use in this research.

2.3.2 Copy Whole Text or Short Phrases

When text-copy is chosen as the task, a subsequent decision has to be made on what type of text to show. Here the researcher has two options - show a normal text (or essay style text) to copy or show several short phrases.

In earlier studies (1910s-1980s) the participants were shown documents such as magazine articles (Gentner, 1982, 1983; Grudin, 1983a), book pages (Hershman and Hillix, 1965; Hiraga et al., 1980), paragraphs from a reading test (Salthouse, 1985), or random words and letters (Fendrick, 1937; Hershman and Hillix, 1965) to copy. These were seen as perfectly suitable tasks for the participants, since the majority of participants were secretaries and typists whose job it was to use the typewriter to create copies of such documents.

However, as computers became more widely available, the typical usage shifted from making neat copies of documents to creative typing, and by and large, the task of copying large bodies of text disappeared. This meant that even if a person was skilled at typing, they might not be used to copying a whole document. Therefore, many researchers now use a set of short, easy to remember phrases (Inhoff, 1991; Butts and Cockburn, 2001; MacKenzie, 2002b). Since the phrases were easy to remember, it was closer to creative typing in that there was reduced effort in reading the words.

Using text that is shorter than a word has been used for measuring particular aspects of typing performance. Presenting one letter at a time for the participant to press has been used in text input studies to measure reaction times (Fendrick, 1937; Hayes et al., 1977; Roussos, 1992). Other studies have incrementally reduced the number of letters that are displayed during typing to see how far ahead expert typists read ahead of what they were typing (Hershman and Hillix, 1965). It is clear that although displaying few letters at a time to participants is useful, its unnatural context in which the participants type mean that it is not suitable for use in the current research.

Using short phrases offers the best of both worlds - it allows the researcher to know what the participant had intended to type, but also mimics the more natural task of writing from scratch as close as possible. It is also less likely that children would get lost when short phrases are shown one by one to them. It is for these reasons that a phrase-copying task with presentation of short phrases was chosen over the other options.

2.3.3 Selection of a Phrase Set

Once the use of phrase set is selected, the next decision to make is which phrase set to use. Studies using phrase sets are usually conducted by showing the participant a short phrase (the presented text - PT), and the participant is asked to enter the phrase into the text input device (the transcribed text - TT). The stream of typing (input stream - IS) is recorded for calculation of speed, accuracy and typing errors later (Butts and Cockburn, 2001; MacKenzie and Soukoreff, 2002b). The error rate of the inputted text is calculated by comparing the intended text with the transcribed text. When measuring these variables the text shown to the participant itself becomes an independent variable and careful considerations must be made to ensure it does not cause any variation in the measurements.

Several different strategies have been taken in creating phrase sets. Some examples are newspaper sentences or sentences emulating a conversation (James and Reischel, 2001), collections of sentences taken from magazines and newspapers such as the Brown corpus created by Kucera and Francis (1967) (Logan, 1982; Lewis, 1999) and phrases containing all the letters of the alphabet (Provins and Glencross, 1968; Read and Horton, 2006). Others use input phrases considered familiar to the user (MacKenzie, 2002b; MacKenzie and Soukoreff, 2003).

Phrase sets are designed to be moderate in length, easy to remember and representative of the target language. However, current phrase sets are designed with adults in mind and their validity with children is untested. One of the most commonly used phrase sets is one created by MacKenzie and Soukoreff (2003). This phrase set, named Text Entry Phrase Set (TEPS) by MacKenzie (2006) contains 500 phrases with no punctuation symbols and only a few instances of upper case characters. The phrases are on average 4.46 words long. The phrase set has been 'used in recent studies with good results' (MacKenzie and Soukoreff, 2003).

Some studies have used these phrase sets on children (Read et al., 2001; Read, 2005; Read and Horton, 2006). However, due to the adult orientated content of these phrase sets, the researchers had to manually select phrases suitable for use with children. As a result of using phrases selectively, the external validity of the experiment may be reduced. There is a need for a child orientated phrase set. Further, to use a phrase set that are not easy to remember or representative of the target language for one participant group (children) would reduce the internal validity of the study dramatically. The author identified three main categories of problematic phrases in TEPS for use with children.

Unsuitable words for children

```
he played a pimp in that movie
make my day you sucker
you are a capitalist pig
```

Some words are Americanised

```
my favorite sport is racketball
vanilla flavored ice cream
```

Words/terms they may not know

```
the dow jones index has risen
sprawling subdivisions are bad
coalition governments never work
```

For typing data collected between children and adults to be comparable, the phrase set must only contain words familiar to young children. Read et al. (2001) constructed short stories from words on the Stage 1 reading list. This was a list of words children aged 6 to 8 years are expected to be familiar with, and so this seems like a sensible approach. Unfortunately, they did not show any example phrases. It is evident that a new child-friendly phrase set is required, one based on reading lists to ensure all words are suitable for use with young children. A new set of phrases collected from children's books was designed and evaluated in Chapter 5.

2.3.4 Gathering Computer and Typing Experience From Participants

In addition to ensuring internal and external validity of the experiment, care has to be taken to ensure that the experiment is repeatable. This means reporting the experiment, including the participants in enough detail so that the same experiment can be repeated by others to refute or support the findings.

Participants can differ in two main ways - their age and the level of skill at the task. It is generally assumed that the older you are, the better you are at doing something. However, computing is different, particularly in the modern age where children are introduced to computing from a young age. It is likely that some young children are more experienced than some adults are.

In addition to gathering demographic data (age, gender, handedness, etc.), it is essential to collect the previous computer and typing experience of participants, to ensure that previous experience is not a factor in the final result. Work by (Book, 1925; Butsch, 1932; Gentner, 1983; Grudin, 1983a; Salthouse, 1984, 1986) show a significant difference in how novice and expert typists make typing errors.

Despite this, in many typing studies, the computer or typing experience of participants are often not reported (Isokoski and Raisamo, 2000; Read et al., 2001; Read and Horton, 2006). Those that do report on computer experience (CE) only broadly categorise, e.g. the skill level as 'novice' and 'experts' (Provins and Glencross, 1968; Grudin, 1983a). The participants were labelled as expert if they were already employed as sectaries or typists (Diehl and Seibel, 1962; Gentner, 1982; Logan, 1982; Gentner, 1983) or have won typewriting competitions (Book, 1925), whilst typists were labelled as 'novice' if they were attending beginner typing classes at the time of the experiment (Gentner, 1983; Grudin, 1983a). In one of the few works investigating typing with children, Joyce and Moxley stated only that the children had 'little or no previous experience using typewriters or computers' (Joyce and Moxley, 1989) with no substantiation.

This assignment of typing skills from employment or educational background was suitable in the time when learning to type was a specialised skill one had to be formally trained at. However, in the modern day, typing is so ubiquitous that people learn to type through everyday use of computers, and training is only offered to those who have difficulties in typing. Therefore, judging typing skills based on qualification or employment is no longer suitable. These measures of typing experience are neither a rigorous measure, nor easy to quantify in children as it is unlikely that they have any formal typing qualifications or even been formally taught to type.

Another measure is the participants' typing speed - words per minute (WPM). Gentner et al. (1988) stated 'a good office typist, typing at 80 words per minute (WPM)'. Here, a word is considered to be five letters long. He also reported that champion typists can type at twice that rate.' MacKenzie and Soukoreff (2002b) define modest touch-typing to be 20 to 40 WPM, touch typists to be 40 to 60 WPM and skilled touch typists to be above 60WPM. The average WPM of each participant group was used to validate the differences between the groups (Gentner, 1983), or simply state the level of typing ability of the participant group (MacNeilage, 1964; Logan, 1982; MacKenzie, 2002b). Hayes and Reeve (1980) and Fendrick (1937) used a scale of WPM to arrange the participants in order. Salthouse (1985) reported the range of WPM (18 to 113WPM) to show the diversity of typing skill amongst his participants, and supported this with number of months employed at a job where the main role was to type.

However, WPM is a measure of performance and is different from computer experience. Performance is the ability to type, and not an indication of the typist's previous experience. WPM equates to a sprint, where you could sprint 100 meters on two consecutive days and get different times, whereas CE is the fact that you train four times a week and is more constant. WPM is also device dependent. Clearly, you are likely to be faster on a familiar keyboard with a familiar layout, whereas much slower on a device with an unfamiliar design. Since it is not possible to ensure that everyone is as familiar with the chosen text input as other participants are, a device independent measure is more desirable.

Frequency of use has been reported by Butts and Cockburn (2001). They report briefly on the frequency of use in the chosen text input task (SMS messages) as 'no experience', 'between one and five texts per week' and 'more than five texts per week'. Although Lewis (1999) reports using a 'background questionnaire' with the participants, the details of the questionnaire were not discussed. Inoff (1991) reported that all participants self-rated themselves as 'fluent' at typing.

Outside typing studies, computer experience has been actively studied, and has been given many definitions (Kay, 1992). In its simplest terms, CE is 'the amount and types of computer skills a person acquires over time' (Howard and Smith, 1986). CE is considered to be multifaceted (Heinssen et al., 1987; Koslowsky et al., 1990) and as a result researchers have tried to break it up into parts to make it easier to discuss and measure. Smith et al. (1999) suggested measures of CE can be grouped into two distinct categories:

- *The Objective Computer Experience* (OCE), relating to the amount of computer use.
- *The Subjective Computer Experience* (SCE), relating to the personal perception of the experience (Jones and Clarke, 1995).

OCEs are a collection of 'observable, direct and/or indirect human-computer interactions which transpire across time' (Smith et al., 1999). Jones and Clarke (1995) suggested OCE measures can be further divided into: *amount of computer use, opportunity to use computers, diversity of experience,* and *sources of information*.

In contrast, SCE is defined as 'a private psychological state, reflecting the thoughts and feelings a person ascribes to some existing computing event' (Farthing, 1992; Eagly and Chaiken, 1993). SCE is a latent process that exists in people's minds but cannot be observed directly. SCEs are measured in terms of *perceived competency, control* and *perceived usefulness* (Igbaria and Chakrabarti, 1990; Todman and Monaghan, 1994).

The most common method of measuring a person's CE is by using survey methods such as questionnaires filled in by the participant. The questions ask the respondents to either rate themselves on their CE or computer skills, or answer questions relating to each area of OCE and SCE. Questionnaires often ask the respondent to answer questions relating to their OCE in a quantitative manner (e.g. 'how many hours a week'), whereas SCE questions are asked with qualitative answers (e.g. 'do you like, dislike, or don't mind using a computer?')

Although many sources list example questions on small portions of CE, there is a lack of works that list all the questions asked in a questionnaire covering the entire spectrum of CE.

Roussos (1992) reports difficulties in ascertaining accurate CE from children. He carried out interviews with 21 6-year old participants, asking them 50 questions. Of particular interest here were the questions he asked regarding CE. He admits that these measures of CE were not very accurate since they were difficult to quantify. He had to make assumptions such as a child that owned their own computer had higher CE than those who had a computer at home that was not their own and in turn, they had more CE than those that did not have a computer at home. He highlighted a need for a numerical measure to indicate children's CE, such as the number of hours spent on the computer, but recognised that this was not so easy to gain from children.

This section showed that a more rigorous method of gathering participant CE and typing experiences should be used in typing studies. Fortunately, CE measurements are available in the form of questionnaires, which can also be adapted to focus on typing. Although few CE questionnaires do exist for children, a better method of quantifying CE and typing data from children is needed. Chapter 7 explores how to measure CE from adult and child participants.

2.3.5 How to Show the Phrases

The first stage of the copy-typing task is to display the text or the phrases to be copied by the participants, called the Presented Text (PT). These are the texts or phrases discussed in Section 2.3.3. In displaying PT, researchers have two choices - on paper or on screen. Here, both methods are analysed for their merits and potential issues with application to children are discussed.

2.3.5.1 Paper-to-Screen Copying

In paper-to-screen copying, a text or phrase to be copied is printed on paper and placed next to the participant's computer, or on cards that are shown individually as the participant types. The participant reads the PT off the paper, and then types the words using the selected text input method. The Transcribed Text (TT) is displayed on the computer screen.

Of the two methods, this method is the easiest in preparation. The researcher selects a text or number of phrases. The order in which they are to be displayed is decided (if shown at once on a sheet of paper), along with the font type and size. The sheets are then printed. If phrases are used, the participants may all receive the same set of phrases in the same order or in different orders, or even different phrases.

The paper-to-screen method is used in many studies (Butsch, 1932; Fendrick, 1937; Hiraga et al., 1980; Gentner, 1982, 1983; Grudin, 1983a; Gentner and Larochelle, 1988; Read and Horton, 2006), particularly in the early days of typewriters where there were no screens. This method has continued to be popular even after the widespread use of the computer because of the ease in which to prepare the material. It is seen as a suitable method to use with adults, since copying text from a piece of paper or a book laid next to the keyboard is a common task. However, it seems unlikely that young children are as used to copy typing from a piece of paper as adults are. If young children are not as adept at paper-to-screen copy-typing as adults are, using this method creates a bias towards the adult participants.

Additionally, the paper-to-screen has a higher Focus of Attention (FOA) than screen-to-screen copying. FOA is the number of attention demands on a task (MacKenzie and Soukoreff, 2002b). For example, in a paper-to-screen copying, the participant must attend to the paper with the text to copy, the keyboard and the screen on which the typing appears - so FOA here is three. Although it can be argued that for expert typing, the FOA is lower (since they do not need to look at the keyboard), majority of young children will not fall into this category.

2.3.5.2 Screen-to-Screen Copying

The second method in displaying PT is to use custom software that displays the text onto the screen. Figure 2 shows a screenshot of one devised by Soukoreff and MacKenzie (2003). The researcher selects a phrase set in advance. The program is set to select which phrases to display, either showing all, some phrases in random order, or particular phrases in a particular or randomised order. Such software also provides a space for displaying what the participant has typed. It also provides a consistent testing environment for all participants, which increases internal validity.

Clearly, placing the PT and TT close together on the screen creates less cognitive load on the participants. The FOA is reduced to two since the participant only needs to attend to the screen and the keyboard - there is a much smaller distance for the eye to travel. These software also display only one phrase at a time, meaning there is no danger of children losing their place like they would on a piece of paper.

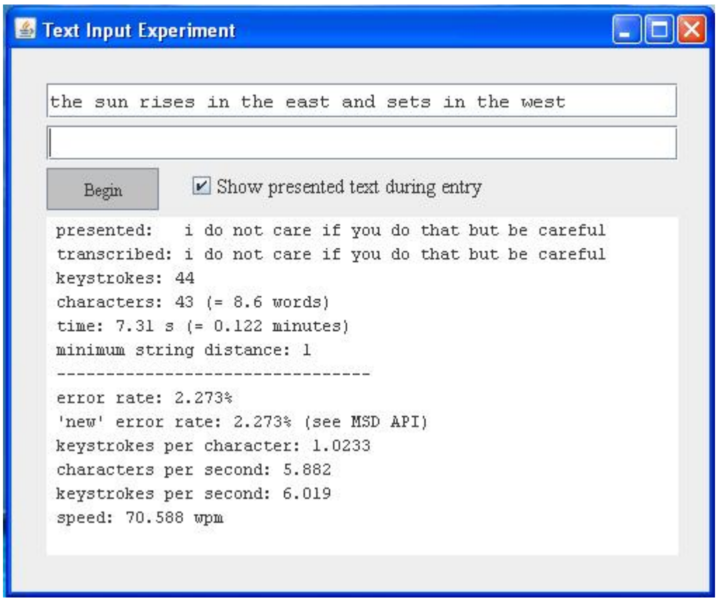


Figure 2: Data Collecting Software by Soukoreff and MacKenzie (2003)

The first disadvantage of this method is the high cost of creating such a program. In studying adults, it is valid to use existing programs described in this section - they have been tested with adults many times. Unfortunately, they remain untested with children. Secondly, in Soukoreff and MacKenzie's (2003) software, once a phrase is typed and the Enter key is pressed, it displays a 9-item statistic summary about the typing. Seeing so many unfamiliar statistics would distract children from the task at hand. It would also make them more aware that they are being tested, and increase anxiety. Thirdly, the font used by Soukoreff and MacKenzie is no larger than font size 12pt. Although this is adequate for adults to complete the task, it may be too small for children.

A second data collecting software, made by Wobbrock and Myers (2006) deals with these issues of distraction and font size. Figure 3 show that their data collector has much less distraction. Besides PT and TT, it only shows the file name to which the data is saved to, whether it is in practice or test mode and how many phrases they have completed. The software has a practice and test mode so that participants typing on a new text input method can practice before carrying out the real test. However, this seems redundant for a study of typing with full-size QWERTY keyboards that the participants are already used to. Additionally, the participants do not need to know what filename their data is saved under, so this also seems redundant for this research.

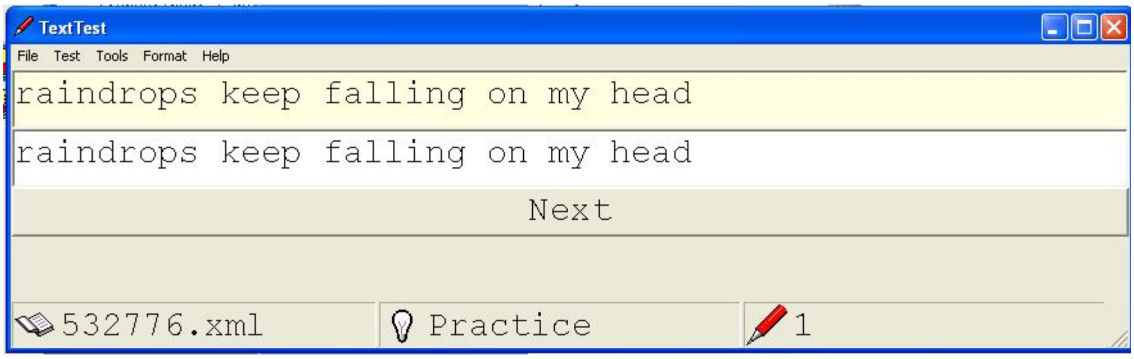


Figure 3: Text Test by Wobbrock and Myers (2006)

The font size is much larger, at approximately 14 or 16pt. This larger font size means it is much easier for young children to read and copy. However, in both data collectors, the font style used has extra decorations on each letter (called Serifs). Opinions on whether Serif or Sans Serif fonts are more readable is divided (Bix, 2002). Many studies have been carried out into the readability of Serif and Sans Serif font styles that have found no significant difference between the two styles (Tinker and Paterson, 1932; Zachrisson, 1965; De Lange et al., 1993) A good review of these studies can be found in Lund (1999). However, since Sans Serif fonts are chosen over Serif fonts by teachers and children's book publishers, children are more familiar with the Sans Serif fonts (Walker and Reynolds, 2002). Therefore, a Sans Serif font style would be more desirable in a data collector for use with children.

Further, these data collectors only gather typing data. If demographic and CE data are also collected from the participants, this has to be done separately, most likely with a paper-based questionnaire. This means the researcher has to maintain two sets of data, one of participant data and one of the typing data. It would be more convenient if the data collector could gather the demographic, CE and typing data all in one, and keep them all on one file.

If the screen-to-screen method is found to be the preferred method for this research, it is imperative that a new data collector (one suitable for children) is produced. The new software would have the following requirements:

- Simplify the interface so there are as few distractions as possible from the typing task
- Use a Sans Serif font style at a font size that is easy for the children to read

- Gather demographic and CE data
- Keep all three (demographic, CE and typing) data on one file

The question of whether the paper-to-screen or screen-to-screen method of presenting the phrases to the participants required further investigations. It is hypothesised that young children would have difficulties in carrying out the paper-to-screen method of phrase-copy typing. If this is the case, a screen-to-screen method will be chosen for this research with a new data collecting software designed for use with children. Chapter 6 attempts to answer this question by evaluating the paper-to-screen method for use with children.

2.3.6 Recording Typing

Two forms of recording what the participant type have been popular in typing analysis studies. First is the use of video cameras to capture visual information of the keyboard as the participant performs the typing. The second is to use key logging software on the computer to unobtrusively record what keys are pressed.

Norman (1983) supported the use of video cameras mounted directly over the keyboard to record the motions of typing. He found it offered an easy frame-by-frame analysis, which revealed the overlapping nature of typing - that movement towards another key is initiated before the previous key has been pressed (Gentner, 1983; Larochelle, 1984). Video recordings also showed that even though each person made consistent patterns of typing, these individual patterns varied markedly between people (Gentner et al., 1980). Grudin (1983a) also used video analysis to determine if a motion towards an omitted key was made (indicating the cause of omission to be not pressing the key hard enough) or not (indicating the cause to cognitive in nature).

Key loggers are software that runs silently in the background on the computer to record the keys being pressed and their timings. The use of the key logger is popular (Gentner, 1982; Logan, 1982; Isokoski and Raisamo, 2000) due to its relative ease to run and analyse the data. It is easy to calculate average length of time taken to press the keys or search for key presses that took longer than a certain amount of time. Automated data collectors such as those described previously in Section 2.3.5.2 also run key logging function in the background to record the keys pressed and their timings.

Video recordings offer a whole wealth of information. Not only do they capture which finger was used to press what key, it also records audio data that is not captured by key loggers. The verbal and non-verbal sounds made by the participants may indicate that they are having some difficulties. Conversations between the participant and other participants or the investigators may also reveal what the difficulty was.

However, videos are an intrusive form of data capture. The video camera pointing at the participants naturally makes them very conscious of being recorded, leading them to behave unnaturally. This is particularly true in children who react to video cameras with a range of behaviours from becoming very static to becoming hyperactive (Markopoulos et al., 2008, p.156). Key loggers on the other hand, can run without the participants feeling so aware of being recorded.

The use of video cameras to record the children's typing is attractive since it is desirable to capture as much data as possible. However, consideration must be made to the effects using video cameras will have on the children. Would the presence of a video camera pointing to their hands as they type be too much of a distraction to young children? If it is found that video cameras do indeed distract the children from their typing task, a key logger, recording their key presses and timings, will need to be used instead. Chapter 6 explores whether children are distracted by the presence of video cameras in a text input task.

2.4 Conclusions

This chapter discussed issues that rise from carrying out typing experiments with children and adults. It showed typing tasks commonly used in studies with adults have several methodological issues that must be addressed before they can be used with children.

2.4.1 Limitations

This review has focused on users of English full-size keyboards. Evaluating reduced keyboards has additional concerns such as whether to use predictive text input method or not, and which predictive method to use. For issues surrounding evaluating mobile devices, readers are directed to (MacKenzie and Soukoreff, 2002b).

Some countries, such as Japan, have native languages with their own alphabets but their keyboard input is based on entering the Roman alphabet. In such cases, teaching children to type is delayed until they have first mastered their own writing system, then taught the Roman alphabet. In such cases, the minimum participant age must be increased, to the age at which the participant is familiar with the characters marked on their keyboard.

The phrase sets reviewed in Section 2.3.3 are all in English. There is an additional consideration for such phrase set for use with children in other languages. For example, in the Japanese educational system children are taught Hiragana, then Katakana, then Kanji. A phrase set designed for young children cannot contain writing systems not yet taught to them.

2.4.2 Contributions

This review has highlighted several issues that must be resolved before a data collection method can be said to produce comparable typing data between children and adults. Firstly, a new phrase set more suitable for use with young children is needed. Secondly, an investigation into whether or not young children are able to carry out paper-to-screen copying efficiently without problems is needed. Thirdly, a new questionnaire is required to systematically collect participants' computer experience with particular focus on their typing. Finally, it must be established as to whether or not the video camera is a suitable method of typing data capture for young children.

2.4.3 Conclusions

Through a review of experimental designs used by other researchers, it was decided that the participants in this study would carry out a phrase-copying task, where they are shown a set of short phrases to type. The copying task was chosen as it is then possible to know approximately what the typist had intended to type. Short phrases were chosen over a long body of text, as they were easier to remember and thus mimic the typing behaviour of creative writing more closely. The next chapter will investigate which, if any, of the existing analysis methods will be suitable for use in analysing children's typing errors.

3 ANALYSING TYPING ERRORS

3.1 INTRODUCTION

Text input experiments generate a set of three strings; the text that is shown to the participant to copy (Presented Text - PT), the string of every key pressed including corrections (Input Stream - IS) and the final text that was generated (Transcribed Text - TT). Once these strings are gathered, the researcher applies analysis methods to extract the desired information. As discussed previously in Section 1.3.2, how these strings are analysed has a crucial role in the outcome of the study. This chapter introduces the wide range of analysis methods available for the purpose of measuring typing errors.

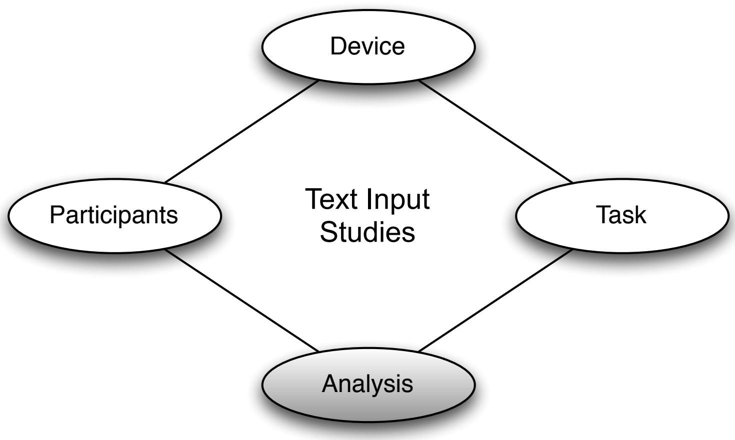


Figure 4: This Chapter Investigates how Typing Data is Analysed in Text Input Studies

There are many ways of measuring how *good* a text input method or a participant is at typing. One approach is to consider a text input to be better if it allows faster input. However, speed alone does not tell how good a participant is at typing. As Soukoreff (2010) highlighted, if a participant typed fast but made many mistakes, it does not mean that they are better at typing than another participant who is slower but makes fewer mistakes.

Researchers investigating text input thus became interested in looking beyond the speed, and looked at the typing errors themselves. Generally, there are two approaches in quantifying typing errors; one is to compute the *error rate*, the other is to *categorise and count the typing errors* found. Both methods provide a way of comparing how good (or bad) devices or participants are at typing. The two methods provide different granularity into the typing errors. Calculating the error rate is a quick method that gives a single figure to provide an overview of how many errors were made per device/participant. In contrast, categorising each typing error provides much richer information on how those errors occurred, but is considerably more time consuming.

In this chapter, both approaches are considered as potential methods of analysis. Requirements for a suitable analysis method for this thesis are defined, and the methods are reviewed according to these criteria. The chapter highlights the need for an evaluation of current analysis methods by applying them to real children's typing data and if required, define a new method.

3.1.1 Objectives

The purpose of this chapter is to introduce methods used to analyse typing errors, to discuss why error rates are not enough for the purpose of this thesis, to introduce how typing errors are classified, to demonstrate that there are many approaches to categorising errors, and to investigate their suitability to the current research.

3.1.2 Scope

As with chapter 2, the investigations carried out in this chapter are limited to those applicable to full-size keyboards. Error rates and categorisation methods specific to other text input methods such as pen and voice inputs are not discussed. For a review of these error metrics readers are directed to a historical review by Soukoreff (2010) and categorical review by Wobbrock (2007).

In describing old typing error categorisation methods, error types only applicable to typewriters (such as 'piling' and 'not inserting the paper correctly') have been removed since they are no longer relevant in modern typing. For an example of error types only applicable to typewriters, readers are referred to Rowe (1931).

3.1.3 Structure

This literature review begins with an introduction to various error rates used in text input research in Section 3.2. It also discuss why error rates do not provide enough information for understanding how two participants differ from each other in the way they make typing mistakes. Sections 3.3 and 3.4 describe existing error type categorisation methods. Section 3.5 demonstrates why a new categorisation method, suitable for use with children, is required. Section 3.6 concludes this chapter with a discussion of the main findings from the literature review.

3.2 ERROR RATES IN TEXT INPUT RESEARCH

An error rate is an aggregated measure of how many errors occurred in a given typing sample. The error rates described in this section each represent a particular facet of the term 'good'. It can represent the ratio of erroneous letters to correct letters typed, how aware the participant is of their own error, or how much error is corrected. It gives a single numerical value for a text input device or a participant that can easily be compared with another. Error rates offer a fast and easy method of comparing two text input methods. Researchers can use one or many error rates to construct their argument for one text input method being 'better' than another.

For the error rates discussed in this section, the below example (Figure 5) is used to demonstrate the calculations. PT is the phrase shown to the participant to copy. IS is the sequence of key presses that the participants made, and TT is the final string produced after the fixes (Soukoreff and MacKenzie, 2001; Wobbrock and Myers, 2006). The Backspace deletes made by the participants are represented by the '<' symbol:

```
PT: the quick brown fox
IS: tha<e p<quik btrwn<<own tox
TT: the quik btrown tox
```

Figure 5: Example of Text Input Data Used for Demonstrating the Error Rates Defined in this Section

3.2.1 Minimum String Distance (MSD)

Minimum String Distance (MSD) measures the accuracy of the transcribed text when compared to the presented text. The Minimum String Distance statistic (also referred to as Edit Distance Algorithm) was first defined by Lowenstein (1966). Soukoreff and MacKenzie (2001) applied this statistic to the comparison of PT and TT. MSD represents the minimum number of primitive edits (single letter omissions, insertions and substitutions) required to convert one string to match the other (Levenshtein, 1966).

Consider PT and TT in the example given in Figure 5. Visual inspection shows that a 'c' was omitted, a 't' was inserted, and 'f' was substituted to 't'. The MSD for this example is therefore three. In other words, the participant must have made at least three character-level errors to produce the TT.

More complex combinations of errors can be hard to compute the MSD for. For this purpose, Soukoreff and MacKenzie (2001) defined an algorithm for computing the MSD of any two pairs of strings. Like the Keystroke Per Character (described later), MSD is therefore easy to compute, and requires no manual calculations.

The MSD compares PT and TT for errors not corrected by the user. By definition, it does not capture any of the errors made and subsequently corrected by the participant. Additionally, what the MSD does not indicate is *what* those uncorrected errors were. It also does not contain any method of representing a word-level error such as an omitted word. Although an omitted word is one error, it is counted as several omitted letters and thus incurs a higher MSD.

3.2.2 Keystroke Per Character (KSPC)

To quantify errors that occurred during typing but were corrected by the participant, MacKenzie (2002a) defined the Keystroke Per Character (KSPC). KSPC is the ratio of the total number of key presses made (every key press in IS, including Backspace) to the final number of characters generated in the transcribed text. KSPC is calculated as:

$$KSPC = \frac{|IS|}{|TT|} \tag{1}$$

A full-size keyboard with a key for each letter has KSPC of 1, but reduced keyboards (such as a 12-key alphanumeric keyboard on a mobile phone) have a higher KSPC. In contrast, the same 12-key keyboard may have a lower than 1 KSPC if predictive text is used. For example, IS in Figure 5 contains 27 characters, and TT contains 19, so KSPC for this typing sample is 1.42. This indicates that for every character in the transcribed text, almost one and a half key presses were made. This formula implies that perfect typing is represented by 1 in KSPC, and the higher this number becomes, the worse the typing. This method is the easiest of the error rates to compute since only the lengths of the two strings are required. Complex comparison between PT, IS and TT is not required.

There are two limitations to this error rate. For the first limitation, consider the following example in Figure 6:

```
PT: the quick brown fox
IS: tha quik btrwn tox
TT: tha quik btrwn tox
```

Figure 6: Same Errors as Figure 5, but with No Corrections

In this case, the same errors were made as that in Figure 5. However, no corrections were made. The KSPC for this example is 18/18 = 1. Even though this example clearly contains several typing errors, the KSPC of 1 implies that the typing was perfect. KSPC penalises efforts made by the participants in correcting the error. The KSPC represents the number of errors made *and* the cost of fixing those errors, without making a distinction between the two. A high KSPC can mean there were many corrected errors, or there were few errors but the cost of fixing those errors was high. Clearly, erroneous letters and the effort of correcting them need to be inspected separately.

Secondly, KSPC offers a measurement of corrected errors by inspecting IS and TT but fails to credit the participant in correcting the errors. MSD measures the uncorrected errors between PT and TT, but does not provide any information on any corrections made. Unfortunately, these two rates are hard to combine to get an over-all picture of the entire typing process. So Soukoreff and MacKenzie (2003) developed a more unified set of error metrics.

In order to calculate the next set of error metrics, each letter in the IS must first be categorised into one of four categories. Each letter found in IS are classified as a Correct (C), Incorrect and Fixed (IF), Incorrect and Not fixed (INF) or Fixes (F). In this taxonomy, the numbers are easy to compute, since Fixes (F) are the number of all the Backspaces found in IS, and IFs are the letters those Backspaces deleted (those letters that appears in IS but not in TT). This leaves all the corrected letters (C) and incorrect errors left in TT (INF). Figure 7 shows how the C, IF, INF and F are counted in the example IS.

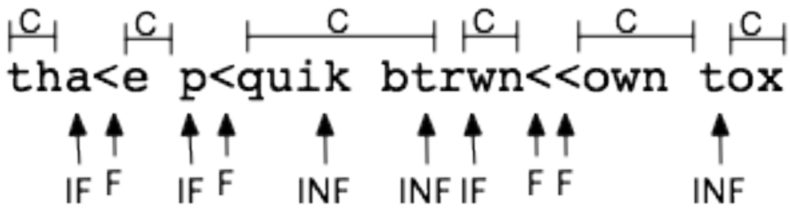


Figure 7: An Example of How to Count C, F, IF and INF From IS

In this example, C = 19, F = 4, IF = 3 and INF = 3. These numbers are to compute the error rates discussed in the following section.

3.2.3 Error Rates (TER, CER and UER)

The first issue this taxonomy corrected was how KSPC did not distinguish between the error and the correction. Here, the distinction is easily made since IF+INF represents all the errors in IS, and F represents all the fixes performed. Soukoreff and MacKenzie (2003) defined the Total Error Rate (TER) as the total number of errors made (both corrected and uncorrected) divided by the total number of letters produced in IS:

$$TER = \frac{IF + INF}{C + INF + IF} \tag{2}$$

Thus the Total Error Rate for the example would be (3+3)/(19+3+3) = 6/25 = 0.24 = 24%. This shows that almost a quarter of all the letters produced were erroneous.

The Corrected Error Rate (CER) is the ratio of incorrect letters that were fixed (IF) to the total number of letters produced:

$$CER = \frac{IF}{C + INF + IF} \tag{3}$$

The Corrected Error Rate for the example is 3/(19+3+3) = 3/25 = 0.12 = 12%.

In contrast, the Uncorrected Error Rate (UER) is the ratio of incorrect letters that were not fixed (INF) to the total number of letters produced:

$$UER = \frac{INF}{C + INF + IF} \tag{4}$$

The Uncorrected Error Rate of the example is 3/(19+3+3) = 3/25 = 0.12 = 12%.

3.2.4 Correction Efficiency (CE)

Soukoreff and MacKenzie (2003) also introduced other useful statistics derived from this new taxonomy. These provide insights in to the efficiency of the participant in the phrase-copying task.

Correction Efficiency (CE) indicates how easy it is for the participant to correct the error. Clearly, the less fixes they have to perform to correct the error, the easier the text input method is to make corrections. Correction Efficiency is defined as:

$$CE = \frac{IF}{F} \tag{5}$$

In the example, Correction Efficiency is 3/4 = 0.75. This indicates that at some point, a letter required more than one Fix to correct the error. Assuming that this was performed on a full-size keyboard (where one key press of the Backspace key deletes one character), it implies that the participant did not notice the error immediately but typed another letter that had to be also deleted.

3.2.5 Participant Conscientiousness (PC)

Soukoreff and MacKenzie (2003) defined Participant Conscientiousness (PC) to represent how many errors made were noticed by the participant then an effort was made to correct them:

$$PC = \frac{IF}{IF + INF} \tag{6}$$

In the example, the Participant Conscientiousness is 3/(3+3) = 0.5 = 50%, so it can be said that the participant noticed and corrected at least 50% of their errors. However, this does not imply that the participant noticed *only* 50% of the errors. It is entirely possible that the participant noticed *all* the errors, but chose to correct only half of them.

3.2.6 Utilised Bandwidth (UB) and Wasted Bandwidth (WB)

The final statistic defined in Soukoreff and MacKenzie (2003) is the Utilised and Wasted Bandwidth. If the act of entering text is viewed as an information transfer from the user to the computer, the Utilised Bandwidth (UB) defines the proportion of bandwidth that transferred useful information.

$$UB = \frac{C}{C + INF + IF + F} \tag{7}$$

Similarly the Wasted Bandwidth (WB) represents the amount of bandwidth that did not result in the transfer of correct letters.

$$WB = \frac{INF + IF + F}{C + INF + IF + F} \tag{8}$$

Wasted Bandwidth can also be calculated as 1 minus the Used Bandwidth. The example calculation is UB = 19/(19+3+3+4) = 19/29 = 0.66, which shows 66% of the bandwidth was used in entering correct letters. In contrast, WB = (3+3+4)/(19+3+3+4) = 10/29 = 0.34, meaning 34% of all key presses produced incorrect letters or fixes.

3.2.7 Why Error Rates are Not Enough

The work by Soukoreff and MacKenzie between 2001 and 2003 (2001; 2002a; 2003) provides a range of error rates that break down the Input Stream into whether the letters were correct or not, and whether the error was corrected. They offer a quick method of comparing different aspects of typing. However, the error rates offer only an overall view of the typing. None discuss *what* the errors were or if there were differences in the cause of these errors. This lack of seeing each error individually means the researcher is no closer to finding out why or how typing errors occur.

Take the example shown in Figure 8. They are two participants in a text input study. They are both presented with the same phrase. They both make some mistakes in copying the phrase.

```
                  PT: the quick brown fox
Participant 1:                      Participant 2:
IS1: the quick thin<<<<brown fox    IS2: than<e quo<ick brt<oww<n fox
C = 19, F = 4, IF = 4, INF = 0      C = 19, F = 4, IF = 4, INF = 0
```

Figure 8: Two Participants Typing the Same Phrase with the Same Error Rates but Completely Different Errors

Although the two participants clearly made very different errors, since the C, F, IF and INF are identical, they will both have the same set of error rates.

The two participants made very different errors. Participant 1 just made a single error of substituting '`thin`' for '`brown`'. In contrast, Participant 2 made four separate errors - a combination of substitutions and Insertions. The two participants made the same *number* of errors on a character-level, but from very different causes. Evidently, in examining the differences in the causes of such typing errors, a different method of quantifying them is required.

Although both error rates and categorisation of errors have been used in text input research, error rates have been used more often in studies comparing text input methods, whereas categorisation of typing errors has been used more in psychology and studies investigating on the causes of error types. Such categorisation methods list definitions of error types the participants could make.

The remainder of this chapter is a chronological summary review of categorisation methods used to categorise typing errors. Although it is not an exhaustive list of categorisation methods, it highlights the different motivations and approaches used in understanding how typing errors were made. As stated in the scope of this chapter (Section 3.1.2), error types specific to only typewriters have been removed for clarity.

3.3 Categorisation Methods in Typewriter Studies

When the typewriter first became commercially available, there were several variations in keyboard size, shape, and layout. In these early days of typing, error investigation was simply about how many letters were typed correctly and how efficient the typist was. However, as interest increased in understanding why users made mistakes, desiring to improve the keyboard and the teaching of keyboard skills, several studies looked at large samples of transcriptions.

In all the studies discussed here, the definitions given for each error type has been directly quoted from the source. Additionally, error types only applicable to typewriters have been removed since they are no longer relevant in modern typing.

3.3.1 Wells (1916)

Wells (1916) collected typing errors from two professional typists and categorised them into error types shown in Table 1. He also measured the time elapsed before, during and after the mistake was made, to assist him in determining the cause of the error types. This work was cited over two decades later by Dvorak et al. (1936) as a seminal work in understanding the cause of typing errors. He found that when the two typists studied were unaware of being studied, they typed at 9/10th of the speed when they were aware of time pressure. He also found that the typists' accuracy and speed was at its highest around noon.

Table 1: Error Types Defined by Wells (1916)

Error Type	Definition	Examples
Excessive action of beginner typist	The slow errors experienced as a beginner typist engages in hunting reaches on an unfamiliar keyboard	
Copy-reading errors	Substitution of words	`ambition -> admiration`
Effective strokes at wrong keys	The stroke is effective, but its play for position and direction sends your finger to the wrong key	
Time delays of error blocking	Blocking (slowing of typing) caused by the interference in typing	
Substitution		Such as neighbouring key strokes and confused vowels
Omission	Omitting a step in the complete sequence	Such as ones difficult to reach (notably, m and n)
Transposition	Interchange of two strokes within the pattern	`engender -> endenger` `unprejudiced -> unpredijuced`
Transposed doubling		`these -> thses` `tyrannised -> tyrranised`
Insertion		
Dominant-sequence interference	More dominant sequences, such as 'the', 'table', 'power' results in substitution and even addition of strokes	`that-> thet` `spectacle -> spectable`

3.3.2 Hoke (1922)

Hoke (1922) examined typing errors to see if they were more likely to occur on certain combinations of letters than other combinations. Errors for each letter of the alphabet were noted until 100 instances were collected. These errors were categorised by the letter before the error. For example, he found that for the 100 errors involving the letter E, 20 occurred after M, 14 after L, 12 after R, 12 after N, and so on. The study suggested that letter combinations least frequently used were more likely to be typed erroneously. This led Hoke to conclude that typing errors are a result of infrequent practice and not caused by the location of keys.

3.3.3 Book (1925)

Book (1925) studied the errors made by typists from varying levels of competencies. His aim was to understand what errors are made at different stages of learning to type. He took typing errors made by 21 typists from five categories of typing expertise ranging from 'amateur' to 'world champions' and attempted to determine the causes of their errors. The typing errors were formally categorised into 17 error types. The 13 still applicable to modern day typing are listed in Table 2:

Table 2: Error Types defined by Book (1925)

Error Group	Error Type	Definition
Finger location of the keys	**Inaccurate location of key - two keys are pressed down at once**	Two keys are pressed down at once because the key to be struck is inaccurately located, the finger usually landing in between the keys.
	Imperfect location of keys - finger misses the proper key	The finger misses the proper key because of imperfect reach, and usually strikes the key next to the correct one, or the one in the same position in the bank of keys above or below the correct one.
	Substitution of one letter for another	Substitution of one letter for another due to substitution of a wrong finger movement for the correct one.
Controlling the sequence of letter-making movements	**Anticipation of letters**	Substitution of a letter that comes later on in a syllable or word for the correct one or for the entire word. Here the intermediate letter is usually omitted.
	Transposition of letters	Adding the correct letter that has been omitted *after* the anticipated letter has been made.
	Addition or insertion of letters, space-bar strokes and extra words	Superfluous or added strokes
	Anticipation of a syllable	Omission of syllables or

	or word and omitting the correct one	words, and adding the omitted word or phrase afterwards.
	Transposition of syllables, words and phrases	-
Getting the copy	**Deviations from copy**	-
	Omission of words or parts of sentences	-
	Insertion or addition of extra words and phrases	-
	Substitution of a wrong word or phrases for the correct one	-
Not formally listed but found in the text	**False strokes**	The individual letter-making movements may be correctly ordered as to sequence and accurately directed but made so lightly that the key struck fails to register its impression on the paper.

Book makes a clear distinction between similar errors by the point at which the error occurs. For example, although inaccurate location of the key and imperfect location of the key can result in the same error, they are distinct in that the former is due to the mind aiming for the wrong key, where as in the latter, the mind aims for the right key, but the execution of the finger was faulty.

He concluded that typists are more prone to make certain errors at certain points in their training, because their skills are not yet fully developed. He also suggested exercises to develop these lacking skills to reduce errors.

3.3.4 Lessenberry (1928)

Lessenberry (1928) carried out a large scale study of letter substitution that has been often cited in later works (Opfer, 1932; Smith, 1932; Dvorak et al., 1936; Grudin, 1983a). To determine the frequency of certain errors, he collected 60,000 typing errors from approximately 6,000 typed papers sent in from all over the United States. Additionally, he also studied daily works carried out by 34 typing students over a period of seven weeks.

For substitution errors, a Confusion Matrix charted the number of times a particular letter was mistyped for another particular letter. This analysis provided information about how likely a letter would be struck instead of another letter. He found that vowels were often confused with other vowels.

Lessenberry, in reporting these results, admitted that the contextual information is lost when only letters are considered (rather than whole words). This was a major limitation of the work, recognising that context is often the key to understanding the cause of the error. Later, Smith (1932) also demonstrated how such character-level counts are artificial and on their own provide little value, as the majority of typing errors are sequences of errors.

Although his Confusion Matrix is commonly cited as the major contribution of this work, Opfer (1932) also cited a categorisation method Lessenberry applied to his errors (Table 3). Most typing errors were found to be due to mental errors involving one or two inaccurate controls of the fingers. These, he suggested, can be corrected easily with typing drill practices.

Table 3: Error Types Defined by Lessenberry (1928), cited in Opfer (1932)

Error Group	Error Type	Notes
Imperfect location of keys	**Inaccurate reach**	Frequently caused by imperfect alignment of the hand with the keyboard. It is caused also by failure to curve the fingers properly, and failure to direct the stroke.
	Transposition of letters	Usually due to reading ahead
Omission of letters, space strokes, words or phrases	**Omission of letters or space strokes**	Usually caused by unevenness of touch
	Omission of entire words	Usually caused by reading ahead
	Omission of phrases or sentences	Caused by raising the eye from the copy.
Inaccurate manipulation of the typewriter	**Faulty shifting**	
	Inaccurate paragraphing	

3.3.5 Clem (1929)

Clem (1929), a typewriting instructor from Wisconsin defined 15 error types for students and teachers of typewriting (Clem, 1929) which can be seen in Table 4 below. However, it is not stated whether these errors were based on any empirical studies and thus is assumed to have been defined 'from experience' of Clem as a typewriting instructor.

Table 4: Error Types Defined by Clem (1929)

Error Types	Examples
Reaching Errors	As string `3` for `e` or `t` for `r`

Substitution Error	Use of the wrong finger or hand for the correct one. As `a` for `s` and `e` for `i`. Wrong word for correct one.
Manipulation Error	Faulty shifting, double spacing due to prolonged space stroke.
Machine Errors	Indented margin or irregular spacing between lines
Speed Errors	Pushing so hard for speed that errors like failing to space between the words or raised capitals from hasty shifting result.
Accidental Errors	The finger slips off the right key to another, which is struck.
Ignorance	Error made because the writer did not know it was an error, e.g. wrong spacing after punctuation marks.
Omission of Letters or Spaces	`tht` for `that`, or forgetting to space between words
Addition of Letters or Spaces	Holding space bar for too long, doubling letters.
Transposition of Letters or Words	Interchanging the letters of a word or interchanging words.
Anticipation Error	Lettering the mind run ahead of the writing, so that some letter in a word ahead is written instead of the letter that should be written, or with the faster writer, the anticipation and writing of a word instead of a letter.
Motorization Error	Made through the influence of a motorized vocabulary. `withing` for `within` or `enought` for `enough`.
Inattention Error	The eye moves to the beginning of a new line of copy and drops down a line too far.
Distraction Error	The attention is distracted from the copy, causing an omission or repetition of words.
Mechanics of Writing	Errors due to incorrect capitalisation, paragraphing, punctuation, syllabication, etc.

Clem stated that this categorisation method was not exhaustive by design, since she felt a simpler method was more practical for use by students and teachers in keeping track of the students' typing progress. She also warned that no matter how thorough a categorisation method may seem, it should not be assumed that it covers everything.

3.3.6 Rowe (1931)

Rowe (1931) attempted to define a thorough categorisation method for Non-Letter Errors. His main aim was to provide a full list of difficulties typewriting teachers could expect their students to encounter when learning typewriting. He analysed errors made by 215 typewriting students from seven different classes over three months. He defined Non-Letter Errors to be those errors involving things other than letters.

Out of 11,180 errors made, 2198 were these Non-Letter Errors, which were further categorised into error types by their causes. 52 error types were defined, concerning the manipulation of the typewriter (e.g. 'using space bar instead of carriage release levers'), the correct positions of fingers ('wrong finger on key'), errors in tasks other than typing (e.g. 'does not hold paper correctly to insert') and knowledge of the typewriter ('does not know how to move paper guide".

Rowe argued that since 20% of all errors were these Non-Letter Errors, it was worthy of further studies and attention paid while learning to type.

3.3.7 Opfer (1932)

Opfer (1932) defined her own classification method by analysing typing errors made by 442 pupils of typewriting classes in California. She grouped these errors according to their causes. The error types were then grouped by whether the error was a Letter Error or a Non-Letter Error. Letter Errors were defined as 'mistakes in striking the correct key to make the letter to be written', and were essentially substituted letters. These errors are listed in Table 5.

Table 5: Letter Errors Defined by Opfer (1932)

Error Group	Error Types	Example
Letter Errors	Due to failure to get the copy	
	Due to anticipation of later letters	
Adjacent Letters	By using wrong finger	`s->d, i->o`
	By wrong reach	`n->m, y->u`
	By correct finger but with the wrong hand	`y->t, g->h`
	By striking the letter in the wrong bank of keys	`w->s, c->d`
Non-Adjacent Letter Errors	By using the wrong finger	`o->u, s->f`
	By using the wrong hand but the correct finger	
	By striking the wrong row key	`e->c, t->v`
	By the wrong reach	`s->e, a->o`

Non-Letter Errors were defined as 'errors that are not necessarily mistakes in striking the correct key, but have to do with manipulation of the machine and operating technique' and is shown in Table 6. Non-letter errors were further grouped by whether the error was due to inattention, ignorance, over-speeding or faulty stroking. Each of these non-letter error groups was subdivided even further by whether it was a mental error or an error in the manipulation of the typewriter.

Table 6: Non-Letter Errors Defined by Opfer (1932)

Error Groups	Mental Errors	Manipulation Errors
Non-Letter Errors of Ignorance	**Wrong syllabication** **Hyphen omitted in a compound word**	**Wrong spacing after a period** **Incorrect spacing after a comma** **Improper indentation** **Spacing after a dash** **Incorrect spacing after a semi-colon**
Non-Letter Errors of Inattention	**Lines too short** **Omission of punctuation** **Incorrect word** **Lines omitted** **Period omitted at the end of sentence** **Lines repeated** **Words repeated** **Incorrect punctuation**	**Spacing within words** **Spacing incorrectly between lines** **Failure to shift** **Failure to indent** **Hyphen struck for tabulator**
Non-Letter Errors of Over Speeding	**Words omitted** **Words inserted** **Hyphen omitted between syllables**	**Transposition of letters in words** **Faulty shifting** **Spacing incorrectly between words**
Non-Letter Errors caused by Faulty Stroking	**Strike-overs** **Letters omitted in words** **Spaces omitted between words** **Letters inserted**	**Letters struck too lightly**

In total, 1129 Letter Errors and 791 Non-Letter Errors were found. Opfer noted that to classify these errors, it was necessary to record not just the word where the error occurred, but also the word before and/or after, and sometimes the line above and below. This, she explained, was required to show the associations and manipulations involved between the remaining text and the erroneous word or letter.

Opfer found that for Letter Errors, the most frequent error was substitutions between `'e'` and `'i'`. For Non-Letter Error, the most frequent error was Wrong Syllabication. Each error type was studied in depth, with likely causes concluded from the pattern of errors found. Opfer used these findings to design remedial exercises for each error type.

3.3.8 Schoenleber (1932)

In an effort to create a chart which lists all the common errors a typewriting student could make, Schoenleber (1932) studied both empirically gathered typing errors and error types defined by previous categorisation methods.

Since Schoenleber's chief aim was to create a chart to assist both students and teachers in identifying their typing errors, she first grouped all typing errors into two categories; - those checked by the students, and those for the teachers to check. These were further grouped according to the cause of the errors. Although Schoenleber defined 59 error types, many are no longer applicable. Table 7 lists 27 error types still relevant today.

Table 7: Error Types Defined by Schoenleber (1932)

Error Group	Error Type	Examples
Faulty use of shift key	**Capitalising letter following capital.**	
	Failure to release shift lock for comma and hyphen when writing all capitals.	
	Failure to shift for special characters.	
Other spacing errors	**Fails to space after word**	
	Too few spaces	
	Too many spaces	
	Transposing last letter of word and space	`space->spac e`
Reading copy - Omission	**Leaving a space in place of a letter**	`apple->a ple`
	Omitting a letter	`apple->aple`
	Omitting a syllable	
	Omitting words	
	Omitting phrases	
	Omitting lines	
Reading copy - Additions	**Superfluous or added strokes**	
	Added syllables	
	Added words	
	Added phrases	
	Added lines	
Reading copy - Transpositions	**letters reversed**	`the->hte`
	Words reversed	`of it->it of`
	Phrases reversed	
	Lines reversed	
Reading copy - Substitutions	**Wrong letter**	
	Wrong word	`there->their`
	Wrong prefix or suffix	`un->in`
Errors in capital	**Failure to capitalise**	

letters	**Capitalising wrong letter**	

Schoenleber used the analysis from this study to create a set of remedial typing exercises designed to be used in conjunction with her error type recording chart. These exercises were aimed at reducing the particular typing errors and contained specially designed remedial drills of words and sentences.

3.3.9 White (1932)

Dvorak et al. (1936) placed significant importance on White (1932) for his work in creating typing drills designed to remedy real life typing errors. To assist typists in increasing accuracy with specific typing drills, White analysed 20,623 typing errors on QWERTY keyboards. He categorised these errors into ten error types covering character-level and word-level errors as shown in Table 8, and considered the ratio of each error type to the overall number of errors found.

Table 8: Error Types Found in White (1932)

Error Type	% of the Total Error (total error = 20623)
Substituted Strokes	40%
Omitted Strokes	20%
Spacing	15%
Transposed Strokes	15%
Inserted Strokes	3%
Double Strokes	2%
Capitalisations	2%
Syllable Division	1%
Reseating Words	1%
Omitting Words	1%

Unfortunately, White's text did not give formal definitions of these errors. He used the analysis to better design typing drills, which included the erroneously typed words found in his study.

3.3.10 MacNeilage (1964)

In understanding the serial order model of typing, MacNeilage (1964) tested five female college students who were touch typists, all at average professional speed (30-45 words per minute). Participants typed up rough reports at home, which were analysed for 623 errors. He grouped his error types into Spatial (where the error was related to where the keys were on the keyboard) and Temporal (errors in the order in which the required letters were typed) as shown in Table 9.

Table 9: Error Types Defined by MacNeilage (1964)

Error Group	Error Type	Definition
Spatial Errors	**Horizontal Error**	Typing a letter immediately to the left or right of the correct letter in the same column of the keyboard.
	Vertical Error	Typing a letter immediately above or below the correct letter in the same column of the keyboard.
	Diagonal Error	A letter is typed in a row and column adjacent to that of the correct letter.
Temporal Errors	**Reversal Error**	Two letters next to each other in the correct sequence are reversed in their order.
	Omission Error	One letter in a sentence is left out.
	Equivocal Error	The letter one stroke ahead of the one required in the copy is typed, after which the subject stops.
	Anticipation Error	A letter is typed which is required more than one stroke ahead of the place where it is mistakenly typed.
Miscellaneous Errors	**Interpolation**	A letter apparently quite unrelated to the correct sequence is inserted.
	Phonemic Error	A letter pronounced similarly to that of the correct letter replaces the correct letter in the sequence.
	Type Error	One letter of a word is changed, making it into a word similar to the correct one but meaningless in context.
	Dynamics Error	The letter adjacent in the sequence to a letter that should have been typed twice, is typed twice instead.
Other Errors	**Multiple Classification Error**	Error that can be placed in more than one category.
	Unclassifiable Error	Error that could not be placed into any of the above categories.

MacNeilage was one of the first investigators to count those errors not classifiable by the categorisation method, and those classifiable as more than one error type. He found 70% of the errors were classifiable into a single error type, but 20% were classifiable into more than one error type and 10% were not classifiable at all. This result offers the first glimpse into the incompleteness of categorisation methods that do not define these special cases.

3.3.11 Logan (1999)

Logan (1999) carried out a longitudinal study on one professional typist across two decades. He believed analysing a large number of errors created by one person would increase our understanding of the processes and mechanisms underlying skilled copy typing. He collected 3,000 errors (out of approximately 1.3 million keystrokes) and classified them into 27 error types (Table 10). Logan based his categorisation on Gentner et al. (1983) and MacNeilage's work (1964), and his contribution is regarded as an expansion of these methods. Although this work was published long after computers and word processors became widely available, the typist carried out her work during the 1960s and 1970s on an electronic typewriter and so this study is classified here as a typewriter study.

Table 10: Error Types Defined by Logan (1999)

Error Group	Error Type	Subcategories
Response Errors	**Omission Errors**	Letter Syllable Word Space
	Substitution Errors	Remote Horizontal Vertical (inc. Number Substitution) Homologous Hand Position
	Insertion (Extra Letter) Errors)	**Perseveration Errors:** Immediate Space Bar Separation Character(s) Separation
		Intrusion Errors: Letter Sequence 'Error Habits' Home letter intrusion
Temporal Errors	**Transposition (Antedating Response)**	One Finger Two Fingers Two Hands
	Antedating Response and/or	**Substitution**: Horizontal Vertical Homologous **Insertion** **Omission**: One Finger Two Fingers Two Hands Space Bar
	Interchange	
	Migration	
	Alternation	
	Doubling	

Linguistic Error	**Antedating Response**	
	Perseveration	
	Another Word	
	Spelling	
Miscellaneous	**Unnamed Error**	

Logan introduces some unique terms in his categorisation method. He defined 'remote' in the substitution errors as a substitution that is not horizontal, vertical or homologous. Homologous was first defined by Genter (1983) and thus will be discussed in detail later, but Logan offered his own interpretation of the error type - 'they are perhaps better conceived as mirror image substitution errors with the left-hand side of the keyboard folded over onto the right'.

Perseveration is defined as repeating a recently performed behaviour – such as immediate insertion of a repeated letter (`differ -> difffer`) or simply doubling a letter (`phenomenon -> phhenomenon`). Antedating Response is when the typist types a letter expected later, such as typing the first letter in the next word for the first letter of the current word, or a letter that appears later in the word (`sufficient -> suffcicent`).

Response Errors had the highest frequency of all the error groups. Logan also counted the number of corrected and uncorrected errors and found only 28% of these remained uncorrected in the transcribed text. This seems high for a professional typist, but Logan stressed that she was instructed specifically not to worry about correcting errors.

Logan concluded that typing errors of highly skilled typists have several error factors. The errors were mostly systematic, with a tendency for the same error to occur in the same word.

3.4 WORD PROCESSING STUDIES

The word processor offered new opportunities for the study of text input, as it became easier to record the Input Stream and keystroke timings. Additionally, new error types specific to the electronic keyboard began to appear such as 'Execution Error' (Read et al., 2001).

3.4.1 Gentner et al. (1983)

In 1983, a major work in studying typing was carried out by Cooper et al. (1983b) which reviewed previous works in many aspects of typing. Of particular interest to this thesis are Gentner et al. (1983) and Grudin (1983) discussed here. In the book, Gentner et al. (1983) formally defined and listed nine different error types as summarised in Table 11.

Table 11: Error Types Defined by Gentner et al. (1983)

Error Type	Definition	Examples
Mis-strokes	An error that can be traced to inaccurate motion of the finger, as when one finger strikes two keys simultaneously.	
Transposition	When consecutive letters are switched. Also occurs when space or punctuation that precedes or follows the word is switched. Subcategories: **1F, 2F, 2H.**	1F: `kind -> iknd` 2F `this -> tihs` 2H `the ->teh`
Interchange across I letters	Two non-consecutive letters are switched with I letters intervening (I>0)	`major -> jamor`
Migration across M letters	One letter moves to a new position, with M letters intervening between its correct position and its end position (M>0)	`that -> atht` (M = 2)
Omission	When a letter in a word is left out.	`omit -> mit`
Insertion	An extra letter is inserted into a text. Some insertions can be classified as mis-strokes.	`insert -> ignsert`
Substitution	When the wrong letter is typed in the place of the correct letter. Subcategories: **Column, Row, Homologous, Non-specified.**	`small ->smsll`
Doubling Error	Word containing a repeated letter and the wrong letter is doubled	`school -> scholl` `screen -> scrren`
Alternating Error	When a letter alternates with another but the wrong alternation sequence is produced.	`there -> threr`

Transpositions can be further divided, according to the fingers and hands involved. 1F implies that the two letters were typed by one finger, 2F implies that two different fingers on the same hand typed the two letters, and 2H implies that Transposition occurred across two hands.

This is also the first time there is a clear distinction between Transposition and Interchange. Prior to this work, as seen in many categorisation methods described already, the two terms were used without distinction. Additionally, the number of letters between the correct and actual position for Interchange and Migration was quantified for the first time. The term 'Transposed Doubling Error' defined by Dvorak et al. (1936) was described by Gentner et al. (1983) as an 'Alternating Error', which perhaps better describes the nature of the error.

3.4.2 Grudin (1983)

In the same book (Gentner, 1983), Grudin (1983a) carried out a transcription study involving 6 expert typists and 70 beginner typists and categorised their typing errors into Substitution, Insertion, Omission, Transposition and Other.

In addition to the Column, Row, Homologous and Non-specified Substitution errors, he defined Diagonal Substitutions and studied them in detail. To compare the frequency of substitution and distance between the letters, he applied the 1F, 2F and 2H classifications, that Gentner (1983a) used for Transpositions, between intended and actually typed letters. He found that immediately neighbouring letters were significantly more likely to be substituted for each other than those letter pairs farther apart. He concluded that both were likely to be caused by errors in the control of the typing fingers. He also used video recordings to investigate whether typists intended to press the erroneous key or accidentally hit it.

3.4.3 Read et al. (2001)

Read et al. (2001) carried out an evaluation of four different input methods (a QWERTY keyboard, a mouse, speech recognition, and handwriting recognition) with twelve children between 6 and 10 years. These errors, outlined in Table 12, include errors that would not occur in QWERTY typing activities.

Table 12: Errors Types Defined by Read et al. (2001)

Error Type	Example
Cognition Error	Child misreads a word or cannot distinguish letters
Spelling Error	Child misspells words or mispronounces a word that they know.
Selection Error	Child picks '`I`' for '`i`'.
Construction Error	Child cannot form the letter or word correctly. In handwriting, '`a`' may look like '`d`'. In speech, '`dragon`' becomes '`dwagon`'.
Execution Error	The child presses for too long, fails to click or hits the adjacent character.
Software Induced Error	The software misrecognises the word or character

3.4.4 Read and Horton (2006)

In a study focused on typing, Read and Horton (2006) carried out a text copying task with 18 teenagers between 13 and 14 years, analysing the Input Stream for errors. In this work, errors were categorized into six types. Spelling Errors (SE) are, for example, typing '`chemisry`' instead of '`chemistry`'. Next To errors (NT) are created by pressing a key next to the intended key on the keyboard. Close Errors (CE) are similar to NT errors but the key pressed was diagonally adjacent to the intended key.

Double Characters (DC) errors are, for example, typing '`thinn`' instead of '`thin`'. Space errors (SC) are errors such as typing '`overt he`' instead of '`over the`', and Unknown Errors (U) are those errors for which there are no obvious reasons. NT errors were the most common errors. Read and Horton reported several ambiguities in classifying errors. In line with research in this area, they developed a simple algorithm to determine the cause. For instance, they assumed when a key next to the intended key was pressed, it was an NT error, and similarly with CE errors. This was noted to be problematic as an NT or CE error might have been a 'genuine' spelling mistake, for instance, in spelling 'mangre' for 'manger'.

3.4.5 Wobbrock and Myers (2006)

Wobbrock and Myers (2006) took typing error classification back to its basics by classifying errors into Insertion, Omission and Substitutions only, and whether they were corrected or uncorrected. Wobbrock and Myers defined an Insertion as occurring when a letter appears in TT, but not in PT. Omission occur when a letter appears in PT, but not in TT. Substitution occurs when the corresponding letter in PT and TT does not agree. Examples of these definitions are shown below:

Insertion: `correction -> coerrection`

Omission: `correction -> corection`

Substitution: `correction -> corrwction`

All typing errors can be broken down into these three basic error types and therefore this simple method can capture all typing errors. However, this act of breaking down all error types into individual character-level errors has the same effect as that of Smith's (1932) criticism of Lessenberry's (1928) Confusion Matrix. A major limitation of character-level errors on their own is that they provide minimal information on the context of how the error occurred.

The vast number of categorisation methods outlined in this chapter show that there is no consensus as to which method is the best. Researchers often define a categorisation method to suit what they are investigating. It is unclear as to how best to choose one categorisation method from the numerous ones available for a particular purpose. In the next section, four requirements of a categorisation method for this study are defined.

3.5 SELECTING THE RIGHT CATEGORISATION METHOD

The previous two sections outlined the vast range of categorisation methods available already. Using a categorisation method allows for a far more detailed examination of the differences between two participants than error rate does. Studies such as Book (1925) and Grudin (1983a) showed that categorisation methods can offer insights into how typing errors vary amongst the spectrum of typing skills. It is clear that to study whether or not children and adults make typing mistakes differently, categorising typing errors is the most suitable method.

As there were so many of these methods available, a survey of categorisation methods found in literature was carried out to see which method would be most suitable for use in the current work. The review consisted of text input evaluation papers in the field of psychology and HCI between the years 1910 to 2010. The categorisation method must be for typing errors made on full-size keyboards but it could be for the typewriter or the computer. Papers that only used a previously defined categorisation method were not included, but those that extended upon a previous method were included. The categorisation methods could be purely theoretical or based on empirical typing data.

22 separate categorisation methods were identified in the survey. These were reviewed as to whether or not the method assumed formal typing training; restrained from making assumptions to the causes; classified all errors found; and whether it had been tested on children's typing data.

3.5.1 Assuming All Participants to be Formally Trained in Typing

Many categorisation methods reviewed in this chapter included error types that assumed participants to have had extensive formal touch-typing training. Homologous error is such an error type, defined as a substitution of letters that are in 'mirror-image position on the keyboard with respect to the correct key' (Gentner et al., 1983). They defined that when a participant made a homologous error, they pressed a key with the correct finger in the correct position, but on the wrong hand. This definition assumes that the participant is trained to only use a particular finger of particular hand for each key.

Unfortunately, with today's lack of formal touch-typing training, most people type using a selection of fingers to type the same key. Many use the hunt-and-peck method where they use just one or two fingers from each hand to cover the entire keyboard. Even the modern-day 'touch-typists' who can type without looking at the keyboard are likely to be self-trained. Although they use more fingers, often all of their fingers on both hands, their fingers are not restricted to a particular column of keys. Since such typists have the tendencies to use whatever finger is most convenient, often two or more fingers are used for a particular key during a typing session.

It is clearly inappropriate to use an error type that specifies 'correct finger in the correct position' when such guidelines have been ignored by most modern typists. It is therefore important to ensure that the categorisation method selected for this study does not contain any errors that assume the participants to have had formal and extensive touch-typing training. Some of the other error types that make this outdated assumption are:

- Adjacent letter substitution by using the wrong finger (Opfer, 1932)
- Adjacent letters by correct finger but the wrong hand (Opfer, 1932)
- Inaccurate reach (Lessenberry, 1928)
- Any error types defined as being an error between two fingers or two hands, such as those defined by Logan (1999) and Gentner et al. (1983)

3.5.2 Causal vs. Observational Methods

Since the majority of the categorisation methods aimed to understand how errors occurred, it is not surprising that many error types were defined by the cause of the error, rather than by observable features. Error types such as 'spelling error' (MacNeilage, 1964; Read et al., 2001) automatically imply what the cause of the error was. It is reasonable to assume causes of typing errors in adults, who have been extensively studied by psychologists in the 1980s to understand their typing process. However, a question remains as to whether these causes are applicable to children. Do children suffer from the same psychological causes of a typing error? If so, would children display the same symptomatic typing errors in the same manner as the adult typists do?

Due to the lack of substantial research into causes of typing errors in children, it is important to ensure that the chosen categorisation methods remain objective and not make any guesses as to the cause of the typing errors. Therefore, error types such as 'Dominant-Sequence Interference' (Wells, 1916), 'Letters struck too lightly' (Opfer, 1932), 'Accidental Errors, Errors of Inattention, Errors of Distraction, Errors of Ignorance' (Clem, 1929), 'Syllable Division' (White, 1932), and 'Phonemic Error' (MacNeilage, 1964) could not be included in the categorisation method.

Instead, the categorisation method for this work had to only consist of observable error types - those that are descriptive of what one can see in the Input Stream or the Transcribed Text. Some examples of such observational error types are - 'Inserted Letter', 'Omitted Word', 'Substituted Letter' and even errors such as 'Transposition Error' and 'Doubling Error'.

3.5.3 Selective vs. Exhaustive Methods

There are two approaches to capturing typing errors in a categorisation method. One is to attempt to capture and describe all error types (exhaustive method), and the other is to list only those error types that the researcher is interested in and ignore any other typing error (selective method).

Figure 9 shows the number of error types defined in each of the categorisation methods reviewed. Although most categorisation methods defined less than ten error types, there were a few that defined in excess of 40 error types. These larger methods cluster around the early 1930s when research into developing check sheets of typing errors for use by typing students in identifying their own typing error patterns were prevalent. The works carried out by Opfer (1932), Schoenleber (1932) and Rowe (1931) aimed to provide an exhaustive list of errors that students could record their errors against to see how well they were progressing in learning to type.

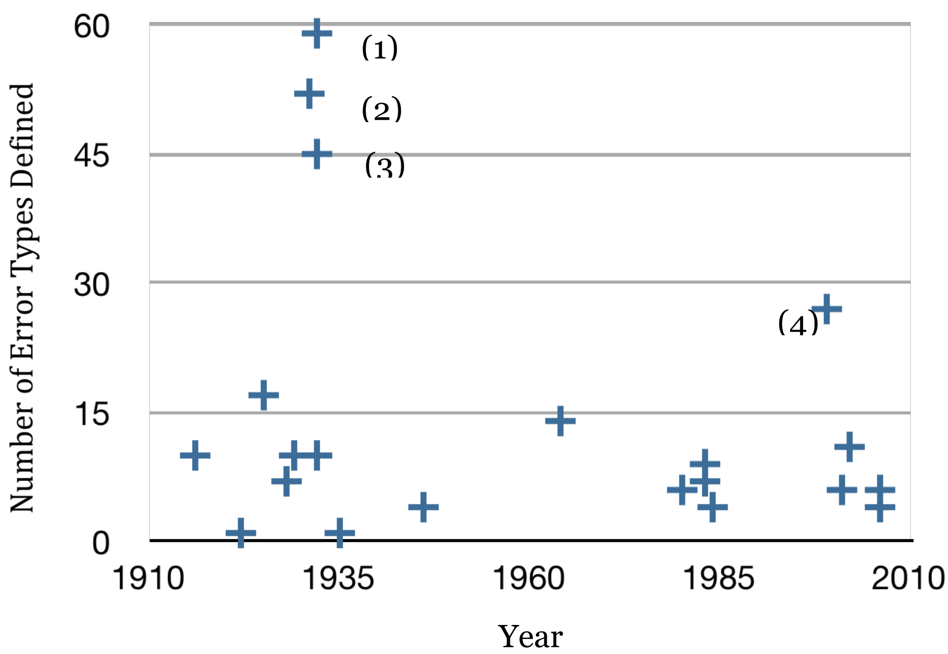


Figure 9: Number of Error Types Defined in Each of the 22 Categorisation Method Surveyed (1 = Opfer (1932), 2 = Rowe (1931), 3 = Schoenleber (1932), 4 = Logan (1999))

The error types they defined were numerous, with (1) Opfer (1932) defining 59 error types, (2) Rowe (1931) 52 error types and (3) Schoenleber (1932) 45 error types. However, Rowe's classification method is not strictly exhaustive. Rowe's study focused on Non-Letter Errors only, such as errors in using the typewriter or in formatting the document. It is possible that Rowe would have defined many more than 52 error types if the focus of the study included errors involving letters.

Unfortunately, trying to use so many error types was cumbersome - it was hard to remember all types - and so was not taken up by later scholars. In addition, a sizeable portion of these error types was specific to the typewriter, such as not having the paper centred correctly in the typewriter (Schoenleber, 1932). These error types became obsolete with the widespread use of the computer.

Later attempts in defining exhaustive categorisation methods became restricted to the typing task only. MacNeilage (1964) and (4) Logan (1999) both took the approach of defining a manageable number of typing errors (14 and 27 respectively) and provide an error type called 'Unclassifiable' to group any typing errors that did not fit into the other error types. MacNeilage also felt that rather than trying to arbitrarily solve errors that could fall into one or more error types, it was better to group these separately as 'Multiple Classification Error'. These exhaustive methods allowed the researchers to gather an overview of all typing errors made. This in turn allowed for understanding of which errors occurred more frequently than others.

In contrast, other researchers focused on particular aspects of typing. In such studies, only a select group of error types were defined. Any typing errors that did not fit into these error types were simply ignored, since they did not offer further insights to the aspect of typing being studied. At the extreme end of these selective methods are works carried out by Hoke (1922). He focused solely on letter substitutions and ignored any other errors. This approach provides in-depth information on a very narrow aspect of typing.

In this thesis, a categorisation method that captures the entire spectrum of typing errors is required. This means that the exhaustive methods are much more suitable for use. However, the method should not have so many error types that using them become cumbersome.

3.5.4 Methods Untested with Children

The final consideration is whether the method is based on empirical data collected from children, or at least has been tested with typing data collected from children. Table 13 shows the age range of the studies reviewed in this survey.

Table 13: Age Range of Participants in the Studies of Categorisation Methods

Author	Age Range (years)
Wells (1916)	Adult
Hoke (1922)	Adult
Book (1925)	Adult
Lessenberry (1928)	Adult
Clem (1929)	n/a
Rowe (1931)	Adult
Opfer (1932)	16-18
Schoenleber (1932)	n/a
White W. T. (1932)	Adult
Davis (1935)	16-21
White M. E (1946)	16-20+
MacNeilage (1964)	20+
Hiraga et al. (1980)	Adult
Gentner et al. (1983)	Adult
Grudin (1983a)	Adult
Grudin (1983b)	Adult
Salthouse (1984)	19-72
Logan (1999)	Adult
Read et al. (2001)	6-10
Berg (2002)	Adult
Read and Horton (2006)	13-14
Wobbrock and Myers (2006)	Adult

Where ages of the participants were not shown, if the participants were described as 'professional typists', an assumption was made that they were at least 18 years old. In studies such as White (1946) and MacNeilage (1964), where only the university year the participants were attending was stated (e.g. 'sophomore college') a guess was made as to the average age of the students. Schoenleber's age range is not shown in this table since he gathered his error types not from typing data but from the teachers that taught typewriting classes.

Although a few categorisation methods were used on teenagers, only one has been used with young children. Unfortunately, this method (Read et al., 2001) was neither exhaustive nor observational, which meant that it was unsuitable for use in this research.

3.5.5 Suitability of Current Categorisation Method for Use with Children

The survey highlighted four important requirements. The chosen categorisation method must not contain error types that assume any formal touch-typing training; only consist of observational error types; the method must be exhaustive; and have been tested with children. Table 14 below chronologically lists all 22 methods and shows which aspects each fulfilled.

Table 14: Categorisation Methods Surveyed Fails to Fulfil All the Requirements

Author	Does not assume formal training?	Observational vs. Causal	Exhaustive vs. Selective	Tested on children?
Wells (1916)	✔	Causal	Selective	
Hoke (1922)	✔	**Observational**	Selective	
Book (1925)	✔	Causal	**Exhaustive**	
Lessenberry (1928)	✔	Causal	Selective	
Clem (1929)	✔	Causal	Selective	
Rowe (1931)		Causal	Selective	
Opfer (1932)		Causal	**Exhaustive**	
Schoenleber (1932)	✔	Causal	**Exhaustive**	
White W. T. (1932)	✔	**Observational**	Selective	
Davis (1935)	✔	**Observational**	Selective	
White M. E (1946)		Causal	Selective	
MacNeilage (1964)	✔	Causal	**Exhaustive**	
Hiraga et al. (1980)	✔	**Observational**	Selective	
Gentner et al. (1983)		Causal	Selective	
Grudin (1983a)		Causal	Selective	
Grudin (1983b)		Causal	Selective	
Salthouwse (1984)	✔	**Observational**	Selective	
Logan (1999)		Causal	**Exhaustive**	
Read et al. (2001)	✔	Causal	Selective	✔
Berg (2002)	✔	Causal	Selective	
Read and Horton (2006)	✔	Causal	Selective	
Wobbrock and Myers (2006)	✔	**Observational**	Selective	

As Table 14 shows, none of the categorisation methods reviewed in this survey is able to fulfil all four requirements. It was clear that there is no one categorisation method that can be used 'as is' for the purpose of this research.

This left two choices for how to go forward - either to take apart and adapt the nearest categorisation method, or to carry out a empirical study of children's typing errors and construct a new categorisation method from the data gathered. A selection of these methods must be tested with real children's typing data before one can be selected or all rejected in favour of a new categorisation method. This is explored further in Chapter 8.

3.6 Conclusions

This chapter has provided a summary of analysis techniques used in text input studies to gauge how 'good' someone's typing is. It showed that although classical error rates are easy to compute, they do not offer the depth of information required in comparing *how* two groups of people differ in the way they make typing mistakes.

A detailed review was carried out on categorisation methods. Section 3.3 and Section 3.4 showed that although there is a vast number of categorisation methods, that there is no consensus as to which one categorisation method is the best to use. Researchers often pick and mix error types that they are interested in and ignore the rest. Conversely, other researchers gather typing data, and then attempt to group all typing errors into something that makes sense. As a result, some categorisation methods contain an impossible number of error types to memorise, and others only offer a small fraction of the whole picture of typing errors.

To select a suitable categorisation method for this study, four requirements were defined. Twenty two existing methods were tested against these requirements but none were found to be completely suitable. Some did not capture all forms of typing errors made, others assumed that the participants had formal and extensive touch-typing training. Most concerning of all, only 1 of the 22 studies was ever tested on typing data collected from children.

This chapter concludes that although many categorisation methods already exist, it is necessary to either adapt an existing method extensively, or to create a new categorisation method that fulfils the requirements identified in this chapter.

3.6.1 Limitations

The literature reviewed in this chapter had to be constrained to those methods applicable to full-size computer keyboards. Many other analysis techniques are available for other forms of text input such as pen, gesture and mobile devices. The current work is focused on the full-size keyboard, and with so many analysis methods available in the literature, an additional deviation into other forms of text input seemed excessive.

In the categorisation methods summarised in this chapter, certain error types were removed for clarity. Errors specific to the typewriter were removed since these were considered obsolete and no longer considered a part of the task of typing.

3.6.2 Contributions

The major contributions of this chapter are the extensive collection made of the typing error categorisation methods and the identification of two new ways to group these categorisation methods (exhaustive vs. selective, causal vs. observational). The chapter also carried out a survey on these methods and has identified a lack of a suitable method that is exhaustive, observational and has been tested on typing data from children.

3.6.3 Conclusions

The analysis methods described in this chapter have been used extensively throughout this thesis. Typing speed and error rates outlined in Section 3.2 are used in Studies described in chapters 5 and 10. Categorisation methods described in Section 3.3 and Section 3.4 form the basis for the new categorisation method defined in Chapter 8.

4 RESEARCH METHODOLOGY

4.1 INTRODUCTION

The aim of the studies in this thesis was to establish differences between how children and adults make typing mistakes. To carry out a valid comparison between these contrasting participant groups, it was crucial to ensure that the data was comparable – that the design and apparatus used had minimal effects on the data. Phrase sets, data collection methods and analysis methods all had to be adapted to be suitable to both children and adults. These adaptations were evaluated by empirical studies in Chapters 5 to 9.

Additionally, although the identification of the cause of any differences observed were not within the scope of this study, it was important to provide a tool that collected enough range of data (particularly computer experience) for other researchers to be able to use it as a starting point of such investigations.

This chapter discusses the decisions made surrounding the research methodologies adopted and their limitations, steps taken to ensure validity and reliability of the data collected, and address the ethical issues in gathering data from children.

4.1.1 Objectives

The aim of this chapter is to discuss in detail the research approach chosen for this study, to highlight the limitations of the selected methods, discuss the validity and reliability of the findings extrapolated from data gathered by such an approach and to address ethical issues regarding collection of experimental data from children.

4.1.2 Structure

Section 4.2 outlines how the research was implemented in terms of collection of data. Section 4.3 discusses the choice of analytical methods used. Section 4.4 discusses the considerations and steps taken to ensure validity and reliability of the findings. Section 4.5 explores the ethical issues surrounding collecting experimental data from young participants. The chapter concludes in Section 4.6 with an overview of the remainder of the thesis.

4.2 RESEARCH DESIGN

This section outlines the main research questions and the research methods used to address them. The discussion also includes the procedures used to improve the internal validity, external validity and reliability of the findings.

4.2.1 The Purpose of Enquiry

The research's main aim was to establish whether children made typing errors in a different way from how adults made typing errors. It focused on typing errors made in phrase-copying tasks carried out on full-size computer keyboards. To do this, a large corpus of comparable typing made by children and adults was required. Chapters 2 and 3 highlighted the need to evaluate current data collection and analysis methods with children and if required, create new methods to achieve comparable data.

Determining the causes of these typing errors was outside the scope of this thesis and therefore, a narrative research where it was possible to ask the participants to say out loud how they think the typing error occurred as they type was viewed as unnecessary. Ethnographic research would provide a wealth of information regarding the environmental causes of some of the error types. This would be useful at a later stage when a researcher may wish to separate error types that occur due to environmental causes (such as distraction from others in the room) from those caused by cognitive processes. Similarly, an interview, perhaps asking the participant to described what they believed happened at each typing error would be useful in determining the cause of each error types. However, since these were beyond the scope of the thesis, ethnography and interviews were also not chosen.

Several research methods have been used on various studies described in this thesis. The details of each method used in these studies are discussed in the relevant chapters. Here an overview of the rationale for the overall approach of this research is presented.

4.2.2 Research Approach Taken

Text input research sits across two closely related but distinct fields - psychology and HCI. The psychologists focused on determining how typing errors occurred, while the HCI field has studied text input methods in terms of measuring usability (or performance) and designing better forms of text input. HCI is also a multi-disciplinary field, receiving major influences from design, science and psychology (Mackay and Fayard, 1997).

In both psychology and HCI, a strong scientific approach is commonly adopted. Both use empirical studies (observations) to refine and prove or disprove a hypothesis, such as whether or not one text input is better than another. Therefore, this research also adopted an empirical approach and closely followed Bryman's (2004) stages of empirical research:

- Choice of research area
- Formulation of research question
- Choice of method
- Formulation of research design and data collection techniques
- Implementation of data collection
- Analysis of data
- Interpretation of data
- Conclusion

With the choice of research area and the research question formulated (Chapter 1), the literature on data collection methods in text input studies (Chapter 2) and typing error categorisation methods (Chapter 3) were studied. This highlighted four areas that threatened the internal validity - phrase set, the collection method, computer experience and the analysis method. These issues required investigation and potentially to be improved before a valid comparison between children and adults could be made.

4.2.2.1 The Phrase Set

The first issue was the phrase set shown to the participants for copying. Current phrase sets were designed with only adult participants in mind. This inevitably meant that the phrase sets suffered from several issues that made them unsuitable for use with children (Section 2.3.3). Using such phrase sets would have reduced the internal validity of any studies involving copying of phrases by both children and adults.

A new phrase set was created with short phrases taken from a selection of children's books (Chapter 5). This new phrase set was compared to the most frequently used phrase set - TEPS (MacKenzie and Soukoreff, 2003) - by the metrics defined in the same paper that measured how close a phrase set was to the natural language. The two phrase sets were also compared through a pilot test with a within-subject repeated-measure design, with participants from two schools. Each participant typed five phrases from each phrase set. Randomisation of the phrase set order reduced the learning effects from affecting the results.

4.2.2.2 Data Collection Method

The second issue was whether to select paper-to-screen or screen-to-screen copying task (Section 2.3.5). Each method had been used in many studies with adults. Adult participants displayed no problem in following a written text on a piece of paper while typing the text out (paper-to-screen). They also displayed no difficulties in copying phrases that appeared on the screen. Therefore, the choice of which method to use has little consequences in adults. However, it was unknown whether children, with less developed cognitive and reading abilities could cope with this method.

The paper-to-screen method was chosen first as it was low-tech and quick to implement. In a pilot study described in Chapter 6, the author used observational techniques to note issues that arose from the paper-to-screen tasks. The conversations between the participants and the investigators, and the typing data was also analysed for any error types specific to this method - i.e. errors that were unlikely to occur if the screen-to-screen method was used. Although observational techniques are highly subjective and likely to alter the participants behaviour (Banister et al., 1998), it was also most suited for highlighting problems that were not obvious from examining the typing log alone.

Due to several issues discovered in this pilot study, the paper-to-screen method was dismissed in favour of the screen-to-screen method. TypingCollector software was created to carry out and gather screen-to-screen phrase-copying tasks, as well as gather demographic and computer experience data from the participants.

The TypingCollector also allowed for a consistent display of the PT and the TT to the participants regardless of the computers used. This increased the internal validity of the research. To ensure measurements were reliable, the TypingCollector recorded the timestamp of each key press, which allowed calculation of the time between key presses accurate to the nearest 1/100th of a second.

4.2.3 Designing Computer and Typing Experience Questionnaires (CTEQs)

Computer experience (CE) was selected as a checking method to ensure that the participant samples were of normal distribution and did not contain deviation from the norm in their previous experience. In designing the CE questionnaire, three survey studies were carried out, one large study (N=137) with undergraduate students and three smaller studies (N = 23, 20 and 48) with children.

The intention of the student study (Section 7.4) was to evaluate which of the many aspects of CE were affected by the sampling method used. Since all participants were first-year undergraduate computing students, it was hypothesised that there would be some positive skew to the distribution of their CE. A survey approach was used to collect a combination of quantitative (e.g. frequency of use and range of software used), and qualitative data (e.g. attitude and self-efficacy of the participants). A test of normality was used to establish which aspects of CE were normally distributed and which were not.

The second study (Section 7.5) investigated two issues surrounding asking children questions relating to their CE. First was whether or not young children understood questions relating to computer hardware and software. Secondly, can children use a Visual Analogue Scale (VAS) to answer questions regarding their CE? A questionnaire was used to collect data for these questions in a within-subject, repeated-measure survey. It asked the children to name images of computer hardware and state what tasks they would use a particular software or hardware for. In evaluating the VASs, in a repeated-measure study, children were asked to answer questions using these new scales. Questions very similar to these were later asked again, but with a completely different format for giving the answers (word clouds).

The third study (Section 7.7) investigated whether or not children were able to answer consistently between paper and computer-based questionnaires. This was so that Child CTEQ (CCTEQ) could be added to the computer-based data collector with confidence that the mode of administration did not have any effect on the answers the children gave. For this, a within-subject repeated measure experimental design was again used, with the children completing both forms of questionnaires in a randomised order. Correlational analysis was made between the answers given on both formats. This study also provided preliminary results on which aspects of CE most correlated with children's typing performance.

The fourth study (Section 7.8) applied the six CE questions defined in the third study to see if there were any deviances from normality in the reported aspects of CE in a sample of children from two schools. A survey approach was used to collect data and a test of normality was used to establish which aspects of CE were normally distributed and which were not.

Two separate CTEQs had to be developed, one for adults and another for children, due to the long and detailed adult questionnaire being too long for children to complete. The TypingCollector displayed the most suitable questionnaire depending on the age of the participant given. This was done so that the typing environment itself remained consistent, ensuring a higher internal validity.

4.2.3.1 The Analysis Method

The fourth issue was in defining a suitable categorisation method. Chapter 3 showed that none of the existent methods fulfilled all of the requirements - observational, thorough and does not assume formal training. Additionally, only one method had been tested with children. In Chapter 8, two existing methods were evaluated with real typing data from children (from the pilot studies described in Chapters 5 and 6) to see how suitable these were.

When the two methods were found to be inadequate for the purpose of this research, they were abandoned in favour of a new categorisation method that achieved all the requirements. The error types were collected from literature and real-life typing data collected in the pilot studies.

4.2.3.2 Automation of Categorisation

Once a suitable categorisation method was defined, it was translated into a Java program (TypingAnalyser) that carried out automatic categorisation. Its main objective was to reduce the time required for analysis of typing errors by accurately categorising typing errors, and reduces ambiguities. An experimental approach was taken to evaluate the TypingAnalyser. A large set of phrase-copying data by children and adults were collected. The data was applied to the TypingAnalyser to check if the analyser categorised all the typing errors correctly. The efficiency of the analyser's disambiguation methods was also tested by comparing the number of ambiguities before and after the disambiguation strategies were applied.

4.2.3.3 Answering the Research Question

With the research design and data collection techniques formulated, it was possible to analyse the typing data from children and adults in an effort to answer the original research question. Using the new data collection and analysis methods, the final study (Chapter 10) analysed typing data collected from 229 adults and 231 children. Where CE was collected, their CE scores were analysed to ensure that CE was normally distributed. Their error rates showed that there were significant differences between the amount of error children and adults made, but that alone did not tell how they differed. Each error type was then analysed error-by-error to discover where the differences lied.

4.3 Ensuring Validity and Reliability

The fundamental consideration that must be made in designing any experiment is whether or not the experimental design employed is valid and that the findings are reliable. A finding from an experiment can only be justified if the experiment was designed in such a way that it had high internal validity, external validity and reliability. In this section, threats specific to this research and actions taken to reduce them are also discussed.

4.3.1 Internal Validity

Internal validity indicates how sure one can be that the independent variable is responsible for the variation in the dependent variables, and not something else that is producing an apparent causal relationship (Bryman, 2004). Blandford et al. (2008) explains that high internal validity means to minimise confounds and make the experiment more robust.

Although selecting a sample from across the country increases the generalisability of the findings (external validity), this meant that there were more variations in the historical factors of the participants, such as education, local dialects and experience. To ensure that the threat of history to the internal validity of the research is kept to a minimum, it was decided that samples would be taken from just the surrounding area of the author's university. This enables the study to be conducted with participants that have all received similar education, and grew up in similar environments.

The threat of variations in instrumentation and administration to internal validity was one of the main reasons why the TypingCollector was created. It enabled all the data collection in as consistent a manner as possible. The TypingCollector asked the questions and gave instructions to each participant in exactly the same way. Every participant saw the text on the screen in the same font style and size.

It is reasonable to think that the older the participants are, the more CE they will have. Indeed, in an ideal experimental design, where CE is not the independent variables, all participants would have exactly the same amount of CE. However, it is not possible to manipulate the amount of computer experience a person has had throughout their life. In such a case, it is crucial to try and accurately measure each participant's CE, so that it can be accounted for. Chapter 7 explores how CE could be measured and what aspects of CE most affect the person's ability to type.

4.3.2 External Validity

The external validity of an experiment questions whether or not the result of a study can be generalised beyond the specific research context (Bryman, 2004). In other words, how applicable is the finding to a different set of people, or place or other conditions. If the result found is only applicable to the selected sample, it is not externally valid. On the other hand, if the result is applicable to a much larger (but similar) population, it has a higher external validity. The way the participants are selected plays a crucial role in external validity. Cook and

Campbell (1979) define external validity as - 'given that there is probably a causal relationship from construct A to construct B, how generalisable is this relationship across persons, settings and times?'.

It is possible that the interaction of selection and treatment could affect the results, since all participants took part on a voluntary basis. It is acknowledged that it may be the case that the better a person is at typing (or at least have a higher belief in one's typing skills) they may be keener to take part, whilst belief that they are slow or an inaccurate typist may discourage them from taking part. However, it was not possible to reduce this threat, since doing so meant forcing everyone in the room to take part, which was deemed unethical.

To reduce the interaction of history and treatment, all studies were carried out on more or less 'ordinary' days where no major event had occurred that was likely to affect the participant group as a whole. It is acknowledged that personal events surrounding the time of the studies (such as bereavement in the family) could affect a person's performance. However, asking whether or not they have recently experienced a death of a family member or if their parents were going through a divorce was deemed as an invasion of privacy.

The effects of pre-testing were avoided by not testing the participants' typing twice. In contrast, it was acknowledged that the effects of 'being tested' were not possible to control, since it was obvious to the participants from the use of a special data collecting software that they were doing something 'out of the ordinary'. However, the internal validity that the TypingCollector affords was deemed a reasonable trade-off.

Furthermore, issues with previous models of data collecting software were highlighted in Section 2.3.5.2. Soukoreff and MacKenzie's (2003) data collector displayed the participant's accuracy in detail when each phrase was typed - this was seen as making the participant more aware of what was being tested, and also distracted the user from the task. Wobbrock and Myers' (2006) data collector made the participants aware when they were in a 'practice-mode' or 'test-mode', thus making them explicitly aware of when they were being tested. The new TypingCollector was designed to be less obvious to the participants that they were being tested. It did not show any performance data to the participants, nor made explicit distinction between practice and test.

4.3.3 Reliability of Data

An instrument of measurement is considered to be reliable if repeated measures produce the same result (Martin, 2008). In terms of experimental design, reliability is concerned with how repeatable the results of a study are (Bryman, 2004).

The stability of answers to the CTEQs has not been tested, since no repeated tests were carried out. If the data from the CTEQ were to be used for more detailed studies than to establish the range of CE in a participant group, a study taking repeated measures to establish the stability of the answers is crucial. However, since the CTEQ scores were used in this research simply to establish that the participant groups had normally distributed CE scores, this was deemed outside the scope of this thesis.

In manually categorising typing errors, the inter-rater consistency is likely to be threatened where two raters categorise the same typing error as being different error types. In addition, ambiguous typing errors (those that could fall into more than one error type) are frequently resolved in an arbitrary manner (Logan, 1999), dependent on what the researcher thinks the error is. These reasons motivated for an automated program that carries out categorisations according to algorithms as an important tool. Such tool allow for a consistent categorisation that does not suffer from inter-rater inconsistencies.

4.4 IMPLEMENTATION OF RESEARCH

This section discusses some of the decisions made regarding how the research was carried out and their justifications. Although every effort was made to carry out studies of high internal and external validity, in some cases, they had to be compromised due to practical reasons. The impact of such compromises made is also discussed here.

4.4.1 Selection of Participants

A total of 229 adult and 231 child participants took part in the research. An additional 112 children took part in the pilot studies. Several considerations were made in selecting appropriate participants for the studies. Selection of participants is always important since not selecting a representative sample could dramatically decrease the external validity of the findings. Blandford et al. (2008) explains that if a non-representative sample of users is involved in the study, the consequences of this for the findings must be carefully considered. In the studies described in this thesis, whole classes were selected, but no individual participants were selected from the class - instead, everyone in the class was given the opportunity to take part.

In children, the educational system naturally provides three different units of sample size - one classroom, one school and several schools. In studies carried out with large sample sizes across several schools (and in particular if these schools were from various regions of the country), the external validity would be high, yet the variety in cultural/regional background, experiences and knowledge would render the internal validity low. In contrast, if only one classroom from one school was used as the sample, the internal validity would be high, since it is likely that all the children have received the same education, and come from the same area. The experiments could also be run in exactly the same settings. However, the external validity would be low since factors particular to that school (such as a head teacher that is keen on Information Technology rolling out advance training for their pupils) would affect the findings. In this thesis, whenever possible, samples have been taken from at least two schools, so that both the internal and external validity would be less compromised. The schools selected for studies in this thesis were all fairly representative of local primary schools. They were all mixed-sex schools, with each class containing a range of abilities. Although there were a few participants that came from an ethnic minority, all participants spoke and wrote English as their first language. This was perhaps not entirely typical of classrooms in such a multi-cultural country, but was seen as justifiable since the phrases used in the studies were in English.

It is however recognised that since both the adult and child samples were taken from those attending schools/university within Lancashire, that the findings may include behavioural patterns that is particular to this area and that without a comparison to other areas, it is impossible to know the effects this has on the external validity of the findings.

It is also recognised that gathering participants from undergraduate computing students has several validity issues. First is that, like many computing departments in other universities, the majority of students were male. Considerations were made to gather participants from other departments (such as the psychology department) to equalise the number of males to females in the sample, but this added another variable, that the two groups of participants were studying completely different degrees, and were likely to have markedly different educational background and computer experience. The second issue was that the external validity of findings from such a sample was low, since university students represent only a portion of the adult population. However, the ease in which to gather large number of participants was considered to be a reasonable trade off.

4.4.2 Selection of Tools

At the start of the research, many schools still lacked computer rooms containing several matching computers. Therefore, the pilot studies were carried out using tablet PCs connected with a full-size keyboard to act like a PC. During the course of the thesis, computer rooms become a standard feature in every primary school, and so for ease of setting up the experiments, the computers found in these rooms were used instead. With this change in the apparatus, two other changes were made - the task changed from paper-to-screen copying to a screen-to-screen task with the use of custom data collecting software that presented phrases for the participants to copy. The logging of the key presses changed from a key logging software to the data collecting software. Care has been taken in this thesis not to mix data collected from these two different settings, since too many variables were changed for the data to be comparable.

In relying on the computer equipment offered by the school, a new variable became apparent. Each school used different makes and models of computer keyboards and screens. For this, care was taken to ensure that all keyboards used had a British QWERTY Windows PC layout with white writing on black keys. The TypingCollector displayed a consistent font (in size and style) across the various equipments.

The TypingCollector was designed to automatically collect data in the background, with its recording being 'silent' to the participants. This was to ensure that there was no loss of data but in a non-intrusive manner. The pilot study also used video cameras to record which finger was used to press each key, but this was abandoned as video cameras were very intrusive to the task - children were often distracted by the video camera being so close to them, and would not focus on the task (Chapter 6).

Each version of the CTEQ was pilot tested with a small group of people similar to the intended participants. This was to ensure that each question was worded in an understandable manner, and that there were no ambiguities to its meanings.

4.4.3 Timing and Location of Studies

Timing of studies had to be carefully considered, especially with children. Participants may be hungry or tired at certain points in the day. For this, all studies described in this thesis were carried out during a normal school/university day. Studies with students all took place during their assigned weekly lab hours.

Location of studies was also carefully considered. Field experiments occur in real-life settings, such as classrooms. In Child-Computer Interaction, this is considered more desirable, as children are more comfortable in their own surroundings and so more likely to act in their normal manner. In contrast, a laboratory experiment gives the researcher time to create in advance a setting that can be more finely controlled. It also allows for the use of more measuring equipments, often adapted for the particular study. However, unfamiliar environment may be stressful to children or conversely, they may be over-excited by being out on a day-trip, away from school. In either case, this change in setting can cause the children to act in an unnatural manner.

Although studies with students were carried out in various laboratory rooms within the department, it was always the room in which the participants usually had their lab sessions. The laboratory rooms also housed identical keyboards and monitors across the department.

There were practical issues surrounding testing several child participants simultaneously in the same room. Children often distracted each other by talking, and 'raced' each other in their progress through the copying task. This was an undesirable behaviour since the intention of the experiments was to gather typing data in as naturalistic a manner as possible. In these cases, children were gently reminded by the investigator that this was not a race, and that they should try and do it in their normal typing speed. Some children were slow typists and attempted to elicit help from their friends in finding particular keys. One even asked her fast-typing friend to do it for her. The investigator made interventions when these elicitations occurred.

4.4.4 Selection of Task and Data

Another consideration made regarding to the collection of typing data was whether to allow the participants to correct their typing errors or not. In this, there were three choices - allow no correction, allow correction as the participant desires and force correction of all typing errors. Both allowing no correction and forcing the participant to correct all errors were deemed too restrictive and made the task untypical. To keep the external validity high, a conscious decision was made to allow the participants to correct any errors that they wanted, but not force corrections.

However, the TypingCollector was not without restrictions to the task. In the first study carried out using the TypingCollector, it was noticed that children were pressing the 'next' button to skip the current phrase without copying them. To address this, a condition was set for each phrase, that the TT was at least 75% of the length of the PT, before the participant could move on to the next phrase. The length of the text rather than the accuracy of the text was chosen as this ensured that the participant made an attempt at typing, but did not exclude errors such as substituted words.

For the gathering of CE data, the adult and child questionnaires were built into the TypingCollector. It was possible to pursue the use of interview or observational methods to gather CE. However, these methods increase the time taken to collect and analyse data, suffer from inter-rater reliability issues and added to the workload of the investigator running the studies. Using these methods would have meant testing children one at a time. This was undesirable since data from large number of participants were required for this research. It was decided that the experiments were easier to run if every data-gathering task was done by one piece of software and the questionnaire was chosen as the best method available in collecting CE data in such a format.

4.5 ETHICAL ISSUES

Although not crucial for the validity of experiments, ethical issues arising from carrying out research, in particular working with children, is another factor that had to be carefully addressed. The research was designed in such a way that the interests and needs of participants were considered. At the start of the research, ethical clearance was obtained from the University Ethics Committee based on an outline of the empirical work to be carried out and considerations were made to ensure that all work conformed to their ethical guidelines.

Blandford et al. (2008) summarised three important elements of ethical considerations as 'VIP' - Vulnerable participants, Informed consent and Privacy, confidentiality and maintaining trust. Strategies taken to address these elements are discussed here.

4.5.1 Vulnerable Participants

Children clearly are vulnerable participants. To safeguard the children, the author and any other investigators that took part in the studies in an assistive role all obtained police clearance. All studies were carried out with at least two investigators present, and care was taken to ensure that no child was left alone with an adult.

Much higher risk than physical harm is emotional harm. It is possible that a child could find taking part in a study stressful, or even distressing. For each study described in this thesis, extensive considerations were made to ensure that the child participants would be as comfortable as possible. Where possible, the studies were carried out in the children's schools. The tasks were designed to be short and not too taxing for the children. They were also reminded at the start and during the task that they could choose to quit if they wanted to, without any consequences.

4.5.2 Informed Consent

Children are also vulnerable in another sense – they are less able to judge the situation, and so are at risk of being exploited. Informed consent is regarded as a crucial tool in ensuring that participants do not commit themselves into something they did not want to take part in. Informed consent emphasises the need to give the prospective participants as much information as needed to make an informed decision about whether or not they wish to take part in the study.

At the start of all studies, the researcher explained the purpose of the study, and clearly described the task that they were about to carry out. They were informed that participation was purely on a voluntary basis and that no recourse would occur from choosing not to take part. Although adult participants were able to make an informed consent based on this, further steps had to be taken with children.

One approach was to ask the parents to give informed consent for their child to participate. Although this is not the same as the child giving informed consent, parents are more able to fully comprehend the task. They are also better judges of what task is likely to be unsuitable or upset the child. Parents were given a full description of the task their child was about to take part in before giving consent. However, gaining informed consent from parents does not stand in the place of informed consent from the children themselves. Therefore, care was taken to explain the task to the children also.

It was possible that children felt they had to take part in the study, for fear of being told off or some other recourse. In all studies (including the adult study), it was explained at the start that even if they opt to take part in the study, they were allowed to change their mind at any point, and stop the participation. Again, they were reminded that no recourse would occur from doing this. In the child studies, this was repeated throughout the study. In particular, if a child looked distressed in any way, the researcher checked with the child that they were okay, and reminded the child again that they could stop taking part.

4.5.3 Privacy

To protect the privacy of all participants, no identifiable data was collected. Demographic data collected consisted of age, gender, academic year they are in and their handedness. No identifiable information such as name, date of birth, religion, ethnicity or information regarding their parents was taken.

In the study described in Chapter 6, video cameras were used to record the children's typing. In an effort to avoid recording the children's faces, the cameras were pointed at the keyboards. The cameras were set to record after each participant sat down and was ready to type, and stopped as soon as the task was complete, to minimise recording their faces as they moved or the camera being knocked over. In the same study, some photographs were taken of the experimental set up with the participants in situ. However, all photos were taken from the back of the participants to avoid taking photos of their faces. For these studies additional consent was obtained from both the children and the parents for the use of video and photographic cameras.

4.6 CONCLUSIONS

This chapter has shown that, where possible, care has been taken to design the research methods used to ensure internal and external validity. The participants' ethical and privacy issues were also considered to ensure that taking part in the study did not distress them in any way.

The remainder of this thesis is structured as follows: the data collection method is designed and evaluated in Chapters 5, 6 and 7, and the analysis method is designed and evaluated in Chapter 8, which is then automated in Chapter 9. The new data collection and analysis method is then used to analyse typing data of 231 children and 229 adults in Chapter 10. The thesis concludes in Chapter 11.

5 PHRASE SET FOR USE WITH YOUNG CHILDREN

5.1 INTRODUCTION

It has become the most common method in text input evaluation studies to show a short phrase to the participant, which the participant then input using the text input method being evaluated. Several collections of such phrases are available to be used by the investigators of text input research. However, as Section 2.3.3 showed, these phrase sets have all been designed for use with adults, and contain words that children may not know. If such a phrase set was used to compare the typing between children and adults, the children are disadvantaged from the start since there are many words in the phrase set that they are unfamiliar with. This chapter focuses on minimising the effect of the phrase set (task) to the overall outcome of a text input study.

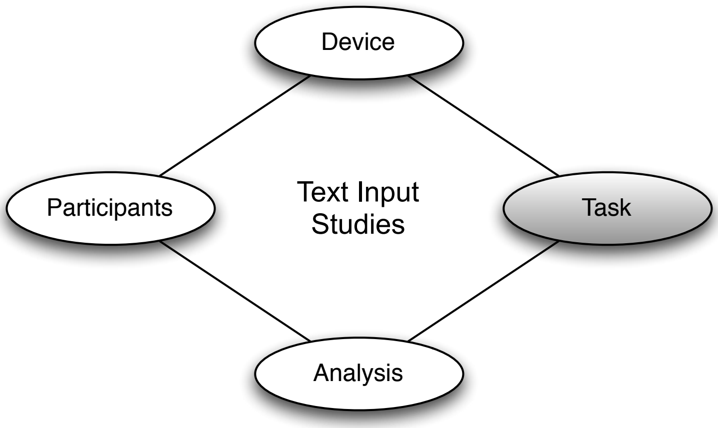


Figure 10: This Chapter Focuses on Minimising the Effect of the Task

A new phrase set was designed to address the issue. The phrases were collected from children's books to ensure that the phrase set contained only words the young participants would be familiar with. This new phrase set allows comparable typing studies between adults and children without reducing the internal validity.

During this study, it was observed that the younger children had difficulties completing the paper-to-screen task. This observation prompted a further, more formal observational study described in the next chapter. The work described in this chapter has been published at NordiCHI 2006 (Kano et al., 2006).

5.1.1 Objectives

The main objective of this study was *to design, create and evaluate a new phrase set suitable for young children*. In addition to this, the study was also a pilot test to observe any obvious issues with children carrying out a paper-to-screen copy-typing task.

5.1.2 Scope

Since the intended use of the phrase sets were on full-size QWERTY keyboards, the study only tested the phrases with this text input method. It is possible that findings may differ with a different text input method.

This study compared two phrase sets by using them in a phrase-copying study with children. The assumption that is being made here is that adults would have no problem in typing the children's phrases, and so a comparison study with adults was not carried out.

5.1.3 Contributions

1. *A phrase set that is suitable for young children*

   The new phrase set can be used with children in a phrase-copying task. The set contains only words familiar to them, rather than words that they are unlikely to have encountered. This increases the internal validity of studies into how children make typing mistakes. Additionally, any researchers using this phrase set with children do not have to pick and choose suitable phrases out of the set, as they have done with previous phrase sets.

2. *Comparable testing of children and adults*

   The data gathered using this phrase set would be comparable between the children and adults, without lowering the internal validity of such studies.

In addition, the casual observations made by the author in this study highlighted the difficulties children faced when completing paper-to-screen phrase copying tasks. This motivated a further study into understanding these difficulties, as reported in Chapter 6.

5.1.4 Structure

Section 5.2 describes the properties of the new phrase set. Section 5.3 evaluates the new phrase set against the most commonly used phrase set. The chapter concludes in Section 5.4, which summarises the main findings of this chapter.

5.2 CHILDREN'S PHRASE SET (CPSET)

Chapter 2 (Section 2.3.3) highlighted the need for a phrase set that can be used with young children. The new phrase set, Children's Phrase Set (CPSet) is intended to be similar to TEPS, but has been adapted for use with children. The set contains 500 phrases taken from children's books and nursery rhymes (Cooling, 1998; Schiller and Moore, 2004). In collecting the phrases from the books, short length sentences (three to seven words) were selected that made grammatical sense on their own. The maximum words per phrase were reduced from MacKenzie & Soukoreff's nine words to seven to minimise variance.

Since the books chosen for collecting the phrases from were written to be suitable for six year olds to read, the phrase set is also suitable for anyone above the age of six years. It contains no capital letters except 'I' and 'J' for 'June', and 'S' for 'Saturday', which were included to see if children would capitalise them or not in a later study. There were also no numbers, and no punctuation symbols. It also contains no American or British specific terms. The full phrase set is shown in Appendix 2.

5.3 EVALUATING THE CPSET

In designing text input method evaluations, MacKenzie and Soukoreff (2003) wrote:

> *'Among the desirable properties of experimental research are internal validity and external validity. Internal validity is attained if the effects observed are attributed to controlled variables. External validity means the results are generalizable to other subjects and situations'.*

This implies that the text entry methods or the devices used become the controlled variable and all other factors should be kept at a constant. It was therefore important to ensure that the choice between using TEPS and CPSet did not significantly affect the results of a text input method evaluation; i.e. that choosing to use the new phrase set would not cause the participants to create more or fewer errors.

5.3.1 Analysing the Phrase Sets

An analysis of the two phrase sets (TEPS and CPSet) was conducted using AnalysePhrase.java (MacKenzie and Soukoreff, 2003) and PHANTIM (Kano, 2005) and the results are shown below in Table 15:

Table 15: Analysis of the Two Phrase Sets

	TEPS	CPSet
PHRASE SET		
Number of phrases	500 phrases	500 phrases
Number of words	2713 words	2350 words
Max phrase length	9 words 43 letters	7 words 34 letters
Min phrase length	3 words 16 letters	3 words 12 letters
Average Phrase Length	5.4 words 28.62 letters	4.6 words 22.0 letters
Number of letters	14310 letters	10998 letters
Correlation with English	0.954	0.982
WORDS		
Number of unique words	1164 words	842 words
Max word length	13 letters	11 letters
Min word length	1 letter	1 letter
Average word length	4.46 letters	3.89 letters
Words containing non-letters	0 words	0 words

CPSet contains 363 fewer words and thus 3312 fewer letters than the TEPS. Each phrase in the CPSet tends to be on average shorter (by 0.7 words), and thus there are fewer letters in each phrase. There are also fewer unique words in CPSet (322 words less). Both phrase sets have high correlation with English (using the letter frequencies of Mayzner and Tresselt (1965)), with 0.982 for CPSet and 0.954 for TEPS.

5.3.2 Validating the CPSet for Use with Children

To compare the two phrase sets, a one-day study was carried out involving 40 children from a local primary school. There were 22 boys and 18 girls, aged between 7 and 10 years old. They are identified as group 1 in the participant summary found in Appendix 3. The study was carried out in a quiet room of a school, using four identical black keyboards (PC Line PCL-K350) connected to four identical tablet PCs (RM Tablet PC CE0984) on stands (not used as tablets, but simply used to create a consistent display). Four children individually carried out the test at a time. Three researchers oversaw the entire study, and there were no video or audio recordings. No names were taken, and results were labelled with only numbers.

5.3.3 Design

In this study the children were asked to copy phrases shown to them on paper, into Notepad™, via a standard QWERTY keyboard. It was decided that each child would type in ten phrases, as from previous experience children of this age group tend to lose interest in the task after copying about ten phrases.

50 phrases were chosen from each phrase set by first randomly choosing a number between one and ten, and then selecting every tenth phrase from the phrase set. The children entered five phrases from one set, then five phrases from the other set. The order of which phrase set was shown first to them was randomized to eliminate any learning effects on their performance. The chosen 100 phrases (50 from TEPS and 50 from CPSet) were each used four times in all.

5.3.4 Procedure

Participants were selected by their teachers but guidance was given by the researchers to ensure a representative sample, with respect to age and gender, was used. The children were asked to sit in front of a tablet PC/Keyboard set up and each had the procedure individually explained to them. The children were each given a sheet of paper with the phrases to type in, presented in Arial font, size 20. Children were instructed to copy the phrases printed on the sheet in front of them into the tablet PC using the keyboard, and were told that the trial was not timed, nor was it marked. During the trial, every keystroke was recorded using KGB Keylogger®. Once the child completed the task, he or she left the room and was replaced by another child.

5.3.5 Analysis

Each phrase was tested four times within this study. Afterwards, a count was made of how many INF (Incorrect and Not Fixed), IF (Incorrect and Fixed) and C (Correct) keystrokes were made by each participant for each phrase. Three values, Total Error Rate (TER), Corrected Error Rate (CER) and Not Corrected Error Rate (NCER) (Soukoreff and MacKenzie, 2003). Comparison was then made between the two phrase sets to see if there were any significant differences in the above three values.

5.3.6 Results

Table 16 below shows a summary of the mean TER, CER and NCER for the two phrase sets.

Table 16: Mean TER, CER and NCER for the Two Phrase Sets

	TEPS	CPSet
TER	**0.0654** (sd = 0.0540)	**0.0636** (sd = 0.0638)
CER	**0.0253** (sd = 0.0240)	**0.0230** (sd = 0.0284)
NCER	**0.0400** (sd = 0.0492)	**0.0340** (sd = 0.0542)

The two phrase sets produced very similar results with regard to error rates. The difference in the TER for the two phrase sets was not significant (Paired T-test, $N = 40$, $t = 0.217$, $p > 0.05$). The differences in CER and NCER for the two phrase sets were also not significant (for CER, Paired T-test, $N = 40$, $t = 1.01$, $p = 0.321$ and for NCER, Paired T-test, $N = 40$, $t = 0.875$, $p = 0.387$). Thus, the performance of the participants in the text input task was not significantly affected by which phrase set the phrases they typed came from.

5.3.7 Discussion

The comparison of the error rates of the phrase set reveals that the participants made equal numbers of errors and fixed similar numbers of errors for each phrase set. It was also observed that the more errors the child made on phrases from one phrase set, the more errors he or she made on the phrases from the other phrase set. This suggests that the two phrase sets can be used interchangeably without lowering the internal validity of a text input method evaluation.

The CER and NCER of the phrase sets confirms this finding; comparison of the CERs show that the children made about the same number of corrections for both phrase sets, and the NCERs show that the children left about the same number of errors uncorrected in their typing.

5.4 CONCLUSIONS

The main contribution of this chapter is the new children's phrase set (CPSet) that is suitable for use on both children and adults in text input evaluation studies.

The results show that there were no significant differences in the performance of the participants across the two phrase sets. This indicates that researchers will be able to choose to use the CPSet for text input method evaluations with children, the content of which is more suitable for children, without the choice affecting their results, and thus not lowering the internal validity of their text input method evaluation.

The CPSet is location independent since it contains no Americanised or British terms, and it can also be used by adults; it offers a common phrase set that can be used in a text method evaluation that involves both children and adults.

The CPSet will allow researchers to run comparison studies in typing by both children and adults, while keeping the internal validity of the experiment high. This is crucial in the current work of understanding how children make typing mistakes differently from adults. The fact that both participant groups will be familiar with all the words shown to them will mean that the phrase set itself is not going to cause any differences in the types of error made between the two groups.

5.4.1 Limitations

This study was carried out on a group of 7 to 10 year olds since they were the youngest group of participants likely to take part in a text input error analysis. Although it is assumed that the adult participants will have no difficulty in copy typing these phrases, the phrase set remains untested with adults.

5.4.2 Conclusion

The new phrase set contains only words familiar to children as young as six years old. Since studies for this thesis will focus on children aged seven years and upward, it is reasonable to say that the phrase set only contains words that all English speaking participants will be familiar with. The CPSet is therefore the first phrase set that allows for comparable data of phrase-copy typing to be gathered between adults and children without decreasing the internal validity.

The author noted during data collection that children displayed some difficulties in carrying out the paper-to-screen phrase-copying task. Some children entered multiple spaces consistently, while others wrote in all capital letters. These difficulties are further investigated in the next chapter.

6 EVALUATING THE PAPER-TO-SCREEN METHOD FOR USE WITH CHILDREN

6.1 INTRODUCTION

This chapter describes a follow-up study from that described in Chapter 5. In Chapter 5, questions arose as to the validity of using a paper-to-screen phrase-copying method with young children. In addition, Chapter 2 (Section 2.3.5) raised the issue that, with their under-developed reading and typing ability, children may struggle to complete this task. If so, this would create a bias towards the adult participants, thus lowering the internal validity of any comparison study using this method. This chapter focuses on minimising this bias in the task.

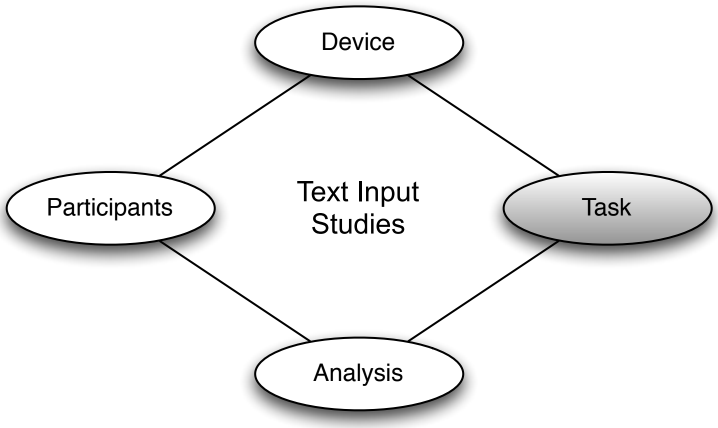


Figure 11: This Chapter Focuses on Minimising the Effect of the Task

The study described in this chapter used the paper-to-screen phrase-copying method with 72 children. Observations made during the study indicate that the paper-to-screen method creates additional difficulties in typing, and so the use of the screen-to-screen copying method is recommended in further studies.

It also found that many children younger than seven years of age did not have enough knowledge about the keyboard to be able to complete the task successfully. In addition, the use of video cameras to record what the participants typed, as done in Grudin (1983a), was too distracting for young children. It is recommended that video cameras are not suitable for use in this manner with children.

The findings from this chapter justified the design of the later studies that used a screen-to-screen phrase-copying task, with children aged above seven years old, with no video cameras.

6.1.1 Objectives

As described above, the primary objective of this study was to answer the question *'does the paper-to-screen phrase-copying method induce its own set of errors in children?'* In addition, the study also answered the following questions:

1. *Is there an age, below which the children are too young to take part in a phrase-copying typing test?*

   Section 6.3 outlines observational evidence that indicate that children younger than seven years did not have enough understanding of the keyboard to carry out typing tasks competently.

2. *Does the use of a video camera in recording what the children typed affect the children's behaviour?*

   Section 6.4.4 argues that the use of video cameras pointing at the keyboard was too distracting for young children. Their attention was focused on the camera and the fact that they were being recorded, rather than on the typing task.

6.1.2 Scope

Since the study's aim was to establish whether or not children had difficulty carrying out the paper-to-screen phrase-copying task, a wider than required age range (5 to 11 year olds) was selected. However, it is acknowledged that they were all recruited from the same school, which limits the external validity of the findings.

6.1.3 Contributions

The main contributions of this chapter are:

1. *Children have difficulties in tracking their place on a sheet of paper when all the phrases are printed on it*

Section 6.4.3 provides observational evidence that children have difficulties in copying phrases from a sheet of paper and suggests that the paper-to-screen method is not suitable for use with young children.

2. *Children aged under seven years are not suitable for phrase copy-typing tasks*

   Section 6.4.1 provides evidence that children younger than seven years of age have difficulties in typing due to a) lack of understanding of the functions on the keyboard, and b) difficulties in translating between the upper and lower case letters.

3. *Set of requirements for a data collecting software for use with young children*

   Section 6.4.5 provides a set of requirements for data-collecting software that will allow for age comparison studies with higher internal validity. Section 6.5 introduces the TypingCollector, which conforms to the new requirements.

6.1.4 Structure

Section 6.2 outlines in detail the experimental study. It describes the method used, the participants selected, hardware and software used, and how the data was analysed. Section 6.3 highlights the results found in the study, implications of which are discussed in Section 6.4. Section 6.4 introduces the new data-collecting program. The chapter concludes in Section 6.6 with a summary of the findings, and a critique of the experimental method used.

6.2 METHOD

In the phrase set study (Chapter 5) the author casually observed some difficulties experienced by children in carrying out the paper-to-screen task. These observations suggested that the method used was causing the children to make errors, which questioned the validity of using such a method. A larger study, carried out over two days, was devised to further investigate these issues. This was an observational study of the same method as described in Chapter 5, with particular interest in the difficulties experienced by the participants.

6.2.1 Participants

72 children, aged between 5 and 10 years took part in this study. 35 were boys and 37 were girls. They are identified as group 2 in the study participant summary found in Appendix 3. These children represented the whole of Years 1, 3, 4 and 5 of the same local primary school. Although all ages between 5 and 10 years were represented, pupils of Year 2 were not available for this study. All children spoke English as their first language, and all children in the classes took part so no selection took place.

6.2.2 Apparatus

The study was carried out in a quiet room of a school, using four identical black keyboards (PC Line PCL-K350) connected to four identical tablet PCs (RM Tablet PC CE0984) on stands (not used as tablets, simply used to create a consistent display). Figure 12 shows this set up.

Figure 12: Set Up of the Experiment

Four children individually carried out the test at a time. Figure 13 below shows that there were two tables in the room, so two children shared a table. To record the typing and the discussions between the participants, a video camera was set up on each table, next to one of the tablet-keyboard set-ups. The cameras pointed directly at the keyboard to avoid recording the participants' faces. It should be noted that only half the children were recorded by video (P2 and P4). Two researchers oversaw the entire study. No names were taken, and results were labelled with only numbers.

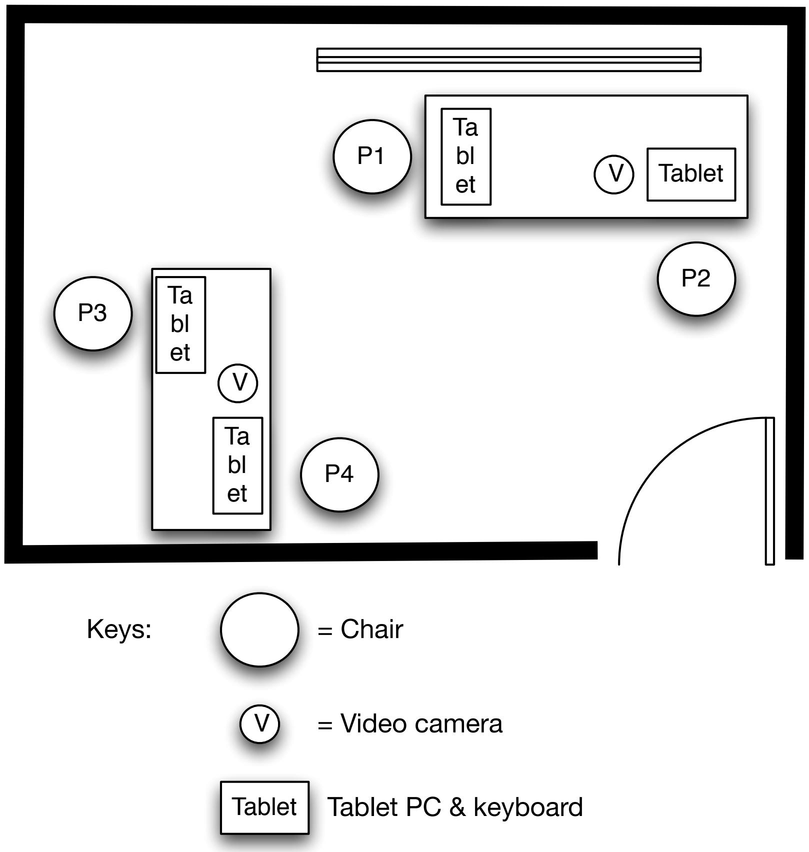


Figure 13: Floor Plan of the Equipment Used in this Study

It can be seen from Figure 13 above that there were only two video cameras available. Therefore, only half of the participants were recorded at the keyboard. However, the video cameras were placed between the two participants on each table in an effort to record what they both said.

6.2.3 Design

This study was a repeat of the previous study (Chapter 5), but with the addition of a video camera to record the conversations between the participants. As before, children were asked to copy phrases shown to them on paper, into Notepad™, via a standard QWERTY keyboard. Consistent with the previous study, the children typed 10 phrases, 5 from the CPSet and 5 from TEPS.

Video cameras were used to record what was typed, and the conversations between the participants. To avoid making the children conscious that their conversations were being recorded and thus restrict what they say, the children were told only that the video camera was there to record their typing. The author made additional observational notes on paper during the study.

Two investigators carried out the study. One was the author, the other was another researcher from the same research group, who assisted in the experiment so that the author had enough time to make observational notes during the experiment. This researcher was briefed about the task before the experiment.

6.2.4 Procedure

The children were asked to sit in front of a tablet PC/keyboard set up and each had the procedure individually explained to them. The children were each given a sheet of paper with the phrases to type in, presented in Arial font, size 20. Children were instructed to copy the phrases printed on the sheet in front of them into the tablet PC using the keyboard, and told that the trial was not timed, nor was it marked. During the trial, every keystroke was recorded using KGB Keylogger®. Once the child completed the task, he or she left the room and was replaced by another child.

Although it is more common to give strict guidelines to the participants about how exactly they should copy PT, it was decided to allow the participants to copy as they felt fit. When a child asked questions regarding the format of TT, such as 'should we use capitals?' the investigators always told the child that they can choose to do it in the way they wanted to. This instruction was given in all the studies reported in this thesis.

6.2.5 Analysis

There were three sets of data collected in this study, the observational notes made by the author, the conversations between the children recorded by the video camera and the key logs of what the children typed. During the experiment, the author made observational notes whenever a difficulty in carrying out the task was noticed. In addition, immediately after the study was completed, the author discussed with the other investigator about if he noticed any difficulties displayed by the children.

6.2.5.1 Observational Notes

The author noted four areas of difficulties during the experiment. These were: difficulties related to the paper-to-screen copying method, problems related to the participant's understanding of the keyboard, issues related to the presence of the video camera, and issues with the format and font of TT.

The younger children in particular displayed difficulties copying with a screen, the keyboard and the PT paper. Year 1 (5 and 6 year old) children were observed to focus mostly on the paper and the keyboard, with little attention given to the resultant text displayed on the screen. This lack of focus on the screen with 6 and 7 year old children has also been observed with pen input in Read et al. (2002). Read et al. also found that the older children (8 and 9 year olds) learnt to check their output on the screen. This was also true in this study, with Year 3 and upwards checking their screens.

In carrying out the typing it was noted that Year 1 children had great difficulty finding letter keys, even after inspecting the keys row by row. The children often called out to themselves the letter they were searching for, and pointed with their finger whereabouts on the keyboard they were searching. This made it easy for the author to see that the children were looking at the correct key, but not recognising it as the key they were looking for. However, when asked by the investigators what the uppercase letter of the intended letter was (e.g. 'what does a big geh look like?'), children were able to find the appropriate key with some searching. This shows that Year 1 children were having difficulties in translating between uppercase and lowercase letters.

Other children, including older children, expressed the wish to start writing on a new line, or delete a letter, but did not know how to do them. They had to have their friends or the investigators show them which key did what they wished to do.

Although Year 1 children showed little interest in how TT looked, older children asked several questions regarding how TT should look like on the screen. They asked whether each phrase should be on a new line, should each phrase start with a capital letter, and whether to use full stops or not.

6.2.5.2 Defining the Coding Scheme for Video Recording

The video cameras recorded the typing behaviour of the participants and the conversations that took place between a participant and another participant, or the participant and the researchers. A transcript was made of all the conversations that were recorded.

The analysis of the video camera purely focused on cases where the participant expressed his or her difficulties to someone else in the room, such as saying 'how do you delete words'. When a participant was seen to be having difficulties but solved the problem by him or herself, or gave up and moved on to another letter, these were not analysed, since doing so required making assumptions regarding their body language alone. Furthermore, any comments that were solicited by the investigators asking the participants questions were removed. This was to ensure that all comments were those unsolicited, thus reducing the influence of the investigators on what the participants said. Notes were also made on comments made regarding the video camera such as 'what is that thing (camera)?'

The four areas identified by the observational notes were used as the codes for a coding of video and typing logs. The items on these logs were categorised into the following:

- Input - problems related to the paper-to-screen copying method (I)
- Output - problems related to the presentation of TT (O)
- Hardware - problems related to the lack familiarity and/or understanding of the keyboard (H)
- Distraction - comments related to the presence of the video camera (D)

By and large, they are mutually exclusive events that could occur in the typing task. Although participants made comments regarding other issues with the task, only these four were used in the coding, since they were already noted as problems. The video analysis was used to confirm these issues.

The codes were applied to each comment that was recorded by the video. Although a small number of comments did not fit in any of the four defined codes, there were no comments that fit into more than one code. The codings of video recordings were carried out only by the author and thus have not been validated by a second observer.

The key logs of the letters typed were manually inspected to see how many phrases the participants completed and if any phrases were omitted. The key logs were also analysed for any typing errors that may be caused by the format of the PT, such as insertion, substitution and omission of words or even phrases.

6.3 Results

The 72 participants copied 633 phrases in total. It became clear early on in the study that Year 1 children (5 to 6 year olds) were not able to type all ten phrases shown to them. The number of phrases shown to them was cut down to five, and a time limit set of 30 minutes. Even with these concessions, some children did not finish typing the phrases. Table 17 shows the average number of phrases typed by each year group.

Table 17: Number of Participants in Each Year Group and How Many Phrases They Copied

	Year 1	Year 3	Year 4	Year 5
Age of participants (years)	5/6	7/8	8/9	9/10
Number of participants	12	20	16	24
Number of phrases completed	51	198	146	238
Average number of phrases per participant	4.25	9.90	9.13	9.92

Due to this uneven number of phrases typed by Year 1, for the remainder of this chapter, frequencies of errors have been normalised by calculating the frequency per phrase per participant.

6.3.1 Video Observations

Table 18 shows difficulties relating to the four areas of interest that were recorded on the video. Numbers in the table indicate the number of times it was observed, per participant per phrase. The errors were logged only if the participant said something to another participant or to the investigators. Things said aloud but to them were not counted.

Table 18: Frequency of each Issue found in the Video Observation (Number of Occurrence per Participant per Phrase)

Code		Year 1	Year 3	Year 4	Year 5
I1	Lost place on sheet	0.0065	0	0	0
O1	Comments regarding the format of TT	0	0.0003	0.0009	0.0009
O2	Comments regarding the font of TT	0.0049	0.0003	0	0
H1	Unable to locate letter key	0.0555	0	0	0
H2	Not knowing a function key	0.0163	0.0010	0.0013	0
D1	Comments regarding the video camera	0.0033	0.0008	0.0013	0.0005

As Table 18 above shows, Year 1 children encountered many more problems than any other group. They required the most assistance in locating the key. However, as noted previously, these children were often able to find the appropriate key when asked what the uppercase letter of the intended letter looked like. Another notable difficulty specific to this age group was losing their place on the sheet of paper, and requiring the assistance of the investigators.

Not knowing what a functional key did was more prevalent. Although this occurred most in Year 1 children, a few children in Year 3 and Year 4 also did not know about the Enter or the Backspace key. However, all children in Year 5 knew how to use these keys. The most troublesome key was the Backspace (8 times), followed by the Enter key (7 times). Other problematic keys were the up arrow, space bar and Caps lock (once each).

The presence of the video camera was noted and commented on by all groups. Children asked why the camera was there and what it was recording. In some instances, a whole discussion about the camera followed.

As noted in the observational note, the older children were interested in getting the format of TT 'right'. The video analysis shows that this increased with the participants' ages. Three comments were made on the font used to display TT. The default font style of notepad was used in displaying TT. Unfortunately, in this default font style (`Lucida Console`), the lowercase letter 'l' looked like an upside-down and back-to-front uppercase 'L'. This caused some confusion in the younger children. Another comment was made that the font size of TT (10pt) was too small.

The video analysis showed that there is an age effect in the comments that are particular to the paper-to-screen method. In the younger children, they have difficulties keeping track of where they are on the sheet. In contrast, the older children were more concerned about matching the visual aspect of PT and TT exactly, such as matching the size of the spaces and replicating the same font style.

6.3.2 Key Log Analysis

In total, the participants copied 633 phrases. In looking at the formatting errors, the focus was on finding participants that had either not understood or misunderstood the act of copying PT. To avoid counting accidental errors, one-off errors were not counted. Only participants who consistently made the same errors were included in the count. Table 19 below shows that the number of participants that displayed these certain traits were counted (and normalised to the number of participants in each year group). This was due to the fact that, if one person consistently adds full stops at the end of all ten phrases, the frequency would be 10, even though only one person did this.

Table 19: Formatting Errors Found (the Numbers Indicate the Normalised Number of Participants that Made the Error per Year Group)

Formatting errors	Year 1	Year 3	Year 4	Year 5
Consistent multiple phrases on one line	0.5	0	0	0
Consistent double spacing	0	0.1	0	0.083
Consistent omission of spaces	0.5	0	0	0

Six children in Year 1 entered the phrase without any spaces between the letters, such as typing '`in the rain`' as '`intherain`'. Some also typed all phrases on one line. This indicates that some Year 1 children do not have the understanding that to copy something means adding spaces as shown in PT, nor to match the layout by typing each phrase on a new line. They perhaps considered typing in just the letters they saw on the sheet into the keyboard was enough.

In addition, two participants consistently added more than one space between each word. A search in the observational notes made during the experiment found that one of them had been noted down for stating that they were putting these extra spaces in to match how big the gaps were between the words in PT. It is assumed that this is likely to have arisen from the fact that PT and TT were shown in different font sizes (10pt and 20pt respectively).

Table 20 shows the typing errors found in the key logs. Here the focus was on those involving entire words or phrases. Omitted Word was the most frequent word-level error observed. It occurred more frequently in Years 3 to 5 than in Year 1. Similarly, inserted, substituted and duplicated words occurred more frequently in the older children than in the younger children. This may be due to the fact that the younger children tend to type slowly, following PT letter by letter. The older children were observed to type several letters or even a few words at a time. The presentation of PT on paper at a considerable distance from TT may have meant that it was difficult for the children to keep track of their place on the paper. It seems reasonable to assume that had the PT and TT been displayed on the same screen, on top of each other, it would have been easier for the children to keep track of what to type next, and also to spot these mistakes.

Table 20: Number of Occurrence per Participant per Phrase of Typing Errors Found that Involved Words and Phrases

	Year 1	Year 3	Year 4	Year 5
Omitted Word	0.003	0.002	0.004	0.001
Inserted Word	0	0.0003	0	0.0004
Substituted Word	0	0.001	0.002	0.001
Duplicated Word	0	0	0.0009	0
Substituted Phrases	0	0.002	0	0
Omitted Phrases	0	0	0.0004	0.0002

Year 1 children made few word level or phrase level errors. During the study, these children were observed reading one letter at a time, and entering it on the keyboard before returning to the paper to read the next letter. The older children (Years 3 and 4) on the other hand were observed reading a word at a time, typing the word, then returning to the paper to read the next word. Although the difference in the unit of reading from PT is likely to be related to the youngest participants not making as many word and phrase level errors, further investigations are required to substantiate this finding.

All occurrences (seven) of Substituted Phrases were carried out by one participant. This participant completed the phrases in the order of 10, 6, 7, 9, 8, 4, and 3. It is assumed that the participant did not see the order in which the phrases must be typed as important. There were two further instances where an entire phrase was omitted.

It is probable that these substitutions and skipping of phrases would have not occurred if the phrases were presented one by one to the children. In designing a suitable method for phrase copying task with children, it is recommended from this finding that PT phrases should be presented to them one at a time. This will reduce the possibility of bias in the data to the adults whom do not commonly experience difficulties tracking which phrase they are typing in a paper-to-screen copying task.

6.4 DISCUSSION

6.4.1 Children Under Seven Years Old are Too Young for the Task

The results from this study give several evidences for the argument that children under the age of seven years are too young to take part effectively in a phrase-copy typing task. It is true that children younger than seven years old are able to press a key to produce a letter on the screen. However, for a comparison study across a range of participant ages to be internally valid, all participants taking part must understand how to type, how to edit and how to format TT correctly so as to match that of PT.

The first problem was the lack of knowledge of the alphabet. Children in Year 1, who so far have spent more time being taught the lowercase letters than the uppercase letters, struggled to type lowercase letters on a keyboard that only showed uppercase letters. 34 instances of Year 1 children unable to find the letter were recorded by the video. Most did manage to find the key for themselves once they asked the question 'what does a big geh look like?'

Secondly, the younger children lacked knowledge of the functional keys on the keyboard. There were two functional keys that every participant used, the Enter key and the Backspace key. Some children did not know how to go to a new line or delete the keys and thus had to ask for assistance.

Finally, the younger children showed signs that they thought to copy a phrase was to simply press the right letter keys. This was indicated by the fact that they failed to insert spaces between words. In addition, some also typed all the phrases on one line.

It is acknowledged here that this study did not include Year 2 children (aged 6 to 7 years) and thus perhaps unfair to state that Year 2 children are too young to effectively take part in typing studies involving phrase-copying. However, some of the difficulties experienced by the Year 1 children were also experienced by some seven year olds in Year 3 (but considerably less frequently). Therefore it is reasonable to say that children younger than seven years old would have significantly more difficulties. It is suggested that for this thesis, the minimum age of participants should be set to seven years old to maintain a higher internal validity.

6.4.2 Paper-to-Screen Method

The results suggest that children had difficulties copying phrases from a sheet of paper to the screen via the keyboard. The key log analysis indicated difficulties were most likely caused by the paper-to-screen method. Some phrases and words were omitted, substituted or inserted in TT. These errors may stem from the fact that there was a considerable physical distance between PT and TT and it was hard to compare if the two matched. It is therefore suggested that the paper-to-screen method should be rejected in favour of the screen-to-screen method. Presenting the PT and TT on the same screen (ideally on top of each other) would make it easier for children to carry out comparisons between the two strings to keep track of where they are on PT, and avoid or spot these errors.

Faced with a sheet of ten phrases, some children had difficulties in keeping track of which phrase they were on. Video analysis showed that children asked for assistance in finding their place on the paper four times. The key log analysis supports this finding in the form of substituted and omitted phrases. To avoid these errors, the screen-to-screen method of presenting PT should only show one phrase at a time.

6.4.3 Presentation of PT and TT

In addition to the paper-to-screen method, there were some issues surrounding the presentation of PT and TT. Some children consistently typed several spaces between each word. This was due to the experiment using Notepad for the participant to type. The default font style (`Lucida Console`) and size (10pt) were used in Notepad, whereas PT was presented in font style Arial, at size 20pt. The difference in font size clearly caused these errors. It is recommended that in future studies the same font style and size should be used for PT and TT.

The study also showed that the font style and size of TT were not suitable. Several children commented on the fact that the lowercase 'l' looked liked an upside-down, back-to-front 'L'. There were also comments suggesting that the children found the small font size hard to work with. It is suggested that a less confusing font in a bigger font size should be chosen for the presentation of PT and TT.

6.4.4 Use of Video Camera

The video camera used for recording the keyboard provided a wealth of quantitative and qualitative information that would have otherwise gone unrecorded. It can record the timing of the key presses, what fingers were used to type what keys, and the conversations that took place during the experiment. However, the presence of the camera was commented upon several times by all age groups. The researcher had to intervene and answer questions about what data the video camera was recording. In one instance, all four children in the room started to have a conversation about the video camera. The presence of the video camera clearly affects the children. If a video camera is not required in a study, it is suggested it is not used. In most cases, the use of a key logging software will suffice.

6.4.5 Custom Data Collector

To fulfil these presentation and recording requirements, it is suggested that custom software is necessary. This study showed that this data collecting software must fulfil the following conditions:

- *Display PT one phrase at a time*

Section 6.3.2 showed that children made errors that were likely to be caused by the fact that all ten phrases were shown on the paper. It is probable that if phrases were displayed one at a time, these word and phrase level errors will be reduced.

- *Display PT and TT on top of each other*

  Section 6.3.1 and 6.3.3 showed that several children lost their place on the phrase sheet due to the large distance between the phrase sheet (PT) and the screen (displaying the TT). It is recommended that PT and TT should be displayed with close proximity to each other.

- *Display PT and TT in the same simple font style and larger font size*

  Section 6.4.3 discussed that older children were concerned about PT and TT being displayed with different font sizes and styles. It is recommended that the two be shown in exactly the same font style and size.

- *Record what key was pressed and its timings*
- *Provide minimum distractions from the task*

6.5 THE TYPINGCOLLECTOR

New data collecting software was written in Visual Basic to gather the demographic and typing data from the participants. The TypingCollector follows the requirements highlighted in the previous section.

The software is designed to automatically run the entire study with minimum interventions from the investigator. This enables a single investigator to run an entire computer lab of participants (approximately 30 participants) simultaneously.

The TypingCollector is designed for Windows-based PCs. It first asks four demographic questions - the participant's age, what academic year they are in, their gender and whether they are right or left handed. This is followed by nineteen questions for the adult participants and six questions for the child participants regarding their computer and typing experience (Chapter 7). Finally, participants are shown ten randomly selected phrases from CPSet, one at a time, to copy.

The phrases were shown in font style Verdana at size point 14, with a space for the participant to copy the phrase underneath. The TypingCollector logged all the answers given, and in the phrase-copying section, what phrases were shown, each key pressed, its timestamp and the time between the keystrokes. Figure 14 shows the TypingCollector phrase-copying stage. This TypingCollector was used for all subsequent studies described in this thesis and provided a consistent environment for the participants.

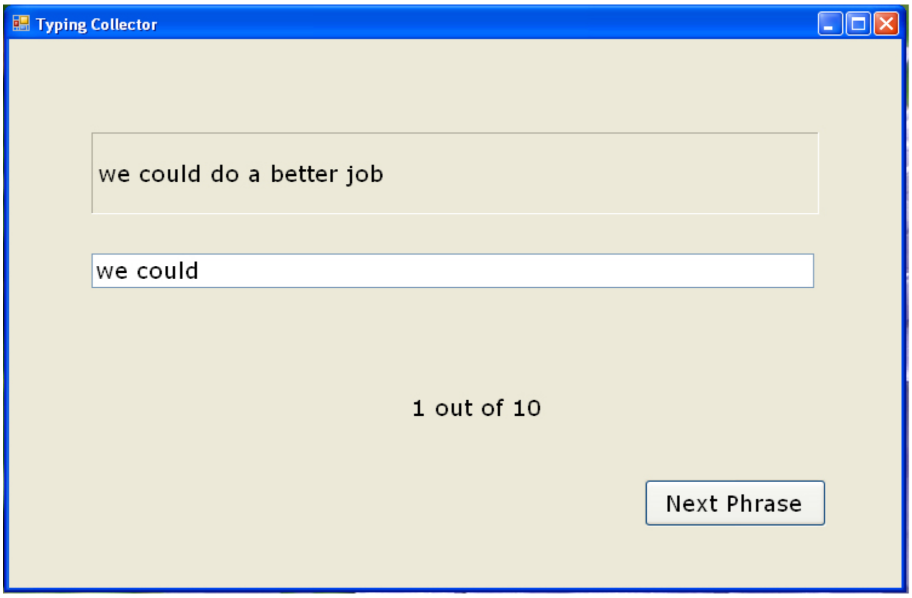


Figure 14: Screen Shot of the TypingCollector Displaying a Phrase with the User Typing Directly Below

6.6 CONCLUSIONS

This study showed that children experience difficulties in the paper-to-screen method in carrying out phrase-copying typing tasks. It concludes that to maintain a higher internal validity when comparing with typing data from adults, the use of a screen-to-screen method would be more suitable. To stop the children from losing track of which phrase they are to type next, the phrases should be shown on screen one at a time.

In addition, the results showed that PT and TT should be presented on the same screen, on top of each other in the same font style and size. This is so that tracking what to type next is easier for the children. A simple font style and larger than 10pt font size should be used in presenting PT and TT.

The use of a video camera close to the keyboard to record the children's typing provides a wealth of information. However, the presence of a video camera distracted the children from the typing task. Therefore, video cameras should only be used when absolutely necessary. Recording what keys were pressed with what timing between the key presses can be done by key logging software, and a the voice recorder may be a less distracting option for recording the conversations.

The findings in this study defined requirements for data collecting software, which were used in designing the new TypingCollector. The TypingCollector was used to gather typing data in all subsequent studies.

6.6.1 Limitations

Although this study had a relatively large participant number (72 children), they all came from just one school. It is possible that if the study was repeated in a different school, with a different curriculum delivery speed, the ages at which different problems manifest may have been altered. However, it is likely that similar sets of difficulties would have been encountered all the same.

It is unfortunate that children from Year 2 (6 to 7 year olds) were not present to take part in this study. The one finding of this study was that seven years old were the minimum suitable age for a phrase-copying task. This was based on the fact that the slightly older seven year olds in Year 3 still had some minor difficulties. However, if a definite minimum age must be established a study of Year 2 pupils will be necessary.

Since only two video cameras were available, only half of the participants were recorded at the keyboard. It is very much likely that some interesting visual observations were missed because of this. However, an effort was made to ensure that the camera sat between two participants so that the conversation between the two could be captured, even if the second person was not visually captured on video. The audio recording on the video was successful at recording even whispers between two participants. One difficulty was that when several people were talking at once, it was hard to decide who was saying what, and so had to be disregarded in the analysis. Fortunately this only happened a few times. Additionally, only the author completed the video analysis. Ideally, a second observer should also carry out the video observation, using the same coding scheme to validate the data.

Finally, an additional evaluator has not validated the reliability of the coding scheme used in this study. By using only four codes to analyse participants' conversations, it is likely that other issues with the paper-to-screen method of copy typing were missed. It is recommended that the coding scheme should be evaluated by other evaluators applying the coding scheme to the data set independently.

6.6.2 Conclusions

The major contribution of the study described in this chapter is that the paper-to-screen phrase-copying typing task is not suitable for children. It also provided several requirements for an improvement to the task method, such as the use of the screen-to-screen method, showing PT phrases one at a time, showing PT and TT on top of each other, in the same simple font style and medium font size, and the use of key loggers with timers rather than video cameras.

A new TypingCollector was created based on these requirements. It is designed to carry out all data collection tasks such as gathering the participants demographic, previous computer experience and typing data. The next chapter focuses on what questions to ask to gather the participant's computer experience.

7 GATHERING PREVIOUS COMPUTER AND TYPING EXPERIENCE DATA

7.1 INTRODUCTION

This chapter focuses on measuring participants' previous computer experience (CE). It should be emphasised that, in this thesis, CE is used as a tool to understand which aspects of CE are affected by the sampling method chosen (such as demographic) and those that are normally distributed within the sample. The investigation of how CE relates to typing performance is outside the scope of this thesis and is not carried out in this chapter.

The participants' CE is a variable within HCI experiments that has a significant effect on computer aptitude (Lee, 1986; Shiue, 2002) For example, a person who has been a programmer for ten years is likely to perform better, even on a brand-new never-seen-before text input method, than a person that has very little CE. To ensure high internal validity in experiments, researchers require a simple method to quantify CE so they can report its effect or design for it.

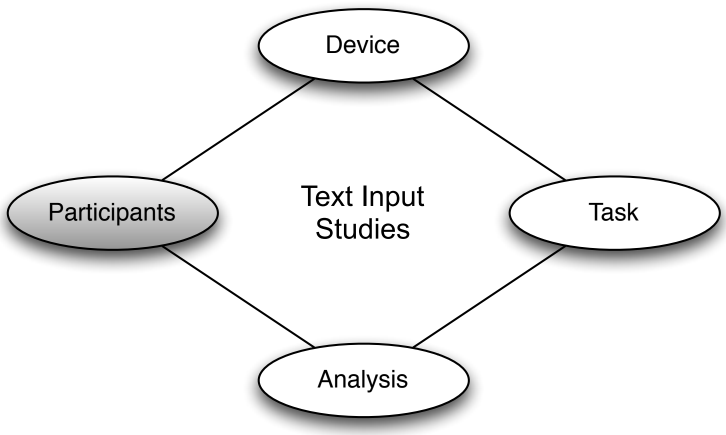


Figure 15: This Chapter focuses on Accounting for the Variation in Participants of the Text Input Study

Most studies of typing skill in adults have classified the participant's CE (for recruitment purposes) by such things as whether they were professional typists or are attending a particular level of typing class (Ford, 1928; Grudin, 1983a; Logan, 1999). However, the use of attendance to typing classes as a measure of CE is neither rigorous, nor easy to quantify in children, since it is unlikely that they have any formal typing training.

Indeed, children's CE is rarely reported in Child Computer Interaction studies. Commonly, only the age of the participants is available. This reduces the repeatability of the reported experiment. Even if you matched the age of the two participant groups, there is no way of knowing if they are matched in previous experience and ability.

This chapter focuses on designing two questionnaires to gather CE data, one from adults and another for children. Both questionnaires were created in a digital format so that the gathering of demographic data, CE data and typing collection could all be done automatically by the TypingCollector. The second study described in this chapter (Section 7.5) was published at NordiCHI 2010 (Kano et al., 2010).

7.1.1 Objectives

The objective of this chapter was to design questionnaires that gathered computer experience data from adults and children. In addition to this primary objective, there were five secondary aims:

1. *Establish the different aspects of computer experience that could provide a detailed overview of an adult participant*

   Section 7.2 surveys the literature on CE and identified 16 aspects of CE. These were transformed into a 19-item questionnaire (Section 7.3) that asked the adult participants questions on the complete spectrum of CE.

2. *Establish whether or not young children understand basic computer hardware and software terms*

   The study described in Section 7.5 showed that young children were able to name basic computer hardware, but were less familiar with software terms. This suggested that it is not suitable to ask young children CE questions regarding software activities.

3. *Establish whether or not children are able to answer questions regarding frequency, attitude, likes and dislike using a VAS*

   In the same study (Section 7.5) two new VASs were tested. The children were able to answer questions using the VAS consistently with other forms of measurements.

4. *Establish whether or not children can complete a computer-based CE questionnaire just as consistently as they do on paper-based questionnaires*

   Section 7.7 showed that children gave similar answers between paper-based and computer-based questionnaires. This suggests that it is acceptable to use a computer-based CE questionnaire for the TypingCollector.

5. *Establish what aspects of CE were affected by the demographic of each participant groups*

   Section 7.4 and Section 7.8 identified that certain aspects of CE were affected by the sampling methods chosen for the study. The adult and child questionnaires provide a method of understanding where sampling bias rise.

7.1.2 Scope

The scope of this chapter is limited to gathering self-reported data regarding the participants' CE. CE measured by observations of their behaviour, or performance, was not considered here. Self-reported data collected through survey methods suffers from its own problems (such as the question of the ability for young children to understand the question asked and give accurate answers). However, it requires the minimum amount of interpretation of the data and thus is less affected by the evaluator's bias.

This chapter also does not include computer anxiety. Although if a participant is highly anxious about using the computer, this may affect their typing ability, it was not considered as part of the person's CE. For meta-analytic review on computer anxiety, readers are directed to (Chua et al., 1999).

Making a single questionnaire for children and adults to extract comparable CE data is a thesis on its own, and clearly beyond the scope of this thesis. Here, the focus was on establishing a CE measure for adults and another for children. Each measure established the range of CE within their respecting participant group and enabled better reporting of CE in studies.

7.1.3 Structure

Section 7.2 first introduces the concept of Objective and Subjective CE, and then describes a literature survey carried out to gather and group CE questions into subgroups. Section 7.3 defines an CE questionnaire for adults, which are tested in a large-scale study carried out with undergraduate students in Section 7.4 to see which of these subgroups were affected by the sampling method chosen. Section 7.5 describes a pilot test conducted with young children (aged 7-10 years old), firstly to see if they understood basic computer hardware terminology, secondly to see if they are able to answer CE questions using VAS. Section 7.6 defines the children's questionnaire. Section 7.7 describes a study that investigated whether or not children were able to answer CE questions consistently between paper and computer-based questionnaires. Section 7.8 investigates the effect of demographics within the selected child sample, before the chapter concluding in Section 7.9.

7.2 MEASURING COMPUTER AND TYPING EXPERIENCE

Section 2.3.4 in Chapter 2 showed how Smith et al. (1999) suggested that measures of CE can be grouped into two distinct categories:

- *The Objective Computer Experience* (OCE), relating to the amount of computer use
- *The Subjective Computer Experience* (SCE), relating to the personal perception of the experience (Jones and Clarke, 1995)

The most common method of measuring a person's CE is by using survey methods such as questionnaires filled in by the participants. The questions ask the participants to either rate themselves on their CE or computer skills, or answer questions relating to each area of OCE and SCE.

Although many literature sources list example questions on select areas of CE, few list all questions that cover the entire spectrum. Others provided overviews of CE but did not list example questions. Additionally, each aspect of CE (such as amount of computer use) appeared to have subgroups within themselves.

To create a CE questionnaire for this thesis, a literature survey gathered example questions used in other CE questionnaires.

Papers discussing questionnaires that contained CE items were studied for any example questions. Each question found was placed in the relevant aspect of CE. These questions were then categorised within each aspect, if necessary, using open coding. In all, 29 papers were used in this survey. The remainder of this section describes the subgroups identified in each aspect, with some example questions that were found.

7.2.1 Objective Computer Experience (OCE)

Objective computer experience (OCE) is the totality of externally observable, direct and/or indirect human-computer interactions which transpire across time (Smith et al., 1999). A person's OCE can be based on their previous and/or current usage of computer technologies (direct OCE), or the medium through which information or knowledge about computers is acquired (indirect OCE) (Jones and Clarke, 1995).

7.2.1.1 OCE1 - Amount of Computer Use

Amount of computer use is the accumulative use of computers (Smith et al., 1999). A survey of the example questions found in literature shows that amount of computer use can be further divided into: when they first started to use a computer, how often they use a computer, how long they use a computer at a time and amount of computer use. Table 21 shows example questions in each of these subgroups. The table also indicates the question code for each question that appears in the questionnaires designed later (see Section 7.3).

Table 21: Subgroups of Amount of Computer Use and Their Example Questions

Subcategories of amount of computer use	Example questions	Question Code (See Section 7.3)
Age first used computer	• The age of first use of a computer (Beckers and Schmidt, 2003)	ACTEQ1 CCTEQ1
Frequency	• On average, how frequently do you use a computer? (Igbaria et al., 1995) • How often it (computer) was used (Robertson et al., 1995)	ACTEQ2 CCTEQ3
Duration	• On an average working day that you use a computer, how much time do you spend on the system? (Igbaria et al., 1995) • Estimated the amount of time each day they spent using a computer at work (Henderson et al., 1995)	ACTEQ3
Amount	• The extent of use as a summation of the hours per application. (Beckers and Schmidt, 2003) • Estimate how many hours per week they had used a	ACTEQ3

	computer over the previous three months (Henderson et al., 1995) • Average usage of applications packages (in hours per week) (Gilroy and Desai, 1986)	

It is generally assumed that the age at which the first encounter with computers took place influences computer experience as well, the younger the age of first use, the more influential this experience may be (Weil et al., 1990). Age of first computer use is of particular importance with young children. It is assumed that children will have increasing CE as they get older, but across an age group, it cannot be assumed that each child will have had the same experience. For example, a 10-year old child that started using computers at 3 years old will have considerably higher CE (7 years) than a child that started using computers at age 6 (CE = 4 years).

However, the age of first use, or the number of years an adult has used computers seems less relevant to their CE. Firstly, it is entirely possible for a computing student to gain more CE by attending few months of their degree, than if someone have used computers for the last 20 years at home but only used it for a limited range of tasks. Secondly, the rate at which computing technology advanced means that the range of tasks that were available to do on a computer 20 years ago was incredibly limited compared to what is available on even the basic entry model of computers these days. This means that a unit of time in using a computer 20 years ago do not equate in terms of CE to the same unit of time using computers today. For these reasons, it is debatable as to whether this subgroup should be included in the questionnaire.

Frequency of computer use is asked in terms of number of times the person uses the computer within a time-span. This time-span should be adjusted depending on the participant group answering the question. For an infrequent user, the time-span may be set at weekly or even monthly frequencies. However, for a group that may all use computers everyday or nearly everyday, this question becomes redundant. In this case, more useful data may be the duration of each session.

Amount of computer use is an informal product of frequency and duration. It can be open to interpretation, and is relative to each participant. For example, a low amount of computer use for a computing student may be considered a high amount of use by a novice user.

7.2.1.2 OCE2 - Opportunity to Use Computer

Opportunity to use a computer is the availability of resources contributing to, or resulting in, the use of computer technologies within or across various settings such as home, school, work (Smith et al., 1999). Opportunity to use computers can be subdivided into ownership, access and training as shown in Table 22:

Table 22: Subgroups of Opportunity to Use Computers and Their Examples

Subcategories of opportunity to use computers	Examples	Question Code (See Section 7.3)
Ownership	• Whether they had a computer at home before attending college (Busch, 1995) • Asked if they owned a home computer (Weil et al., 1990)	ACTEQ4 CCTEQ2
Access	• Computer availability in three areas (computer in the classroom, computers available for teacher use at school, computer available for student use at school) (Rosen and Weil, 1995) • The range of locations in which they (computers) had been used (Todman and Lawrenson, 1992) • Is there a computer in your house? (Roussos, 1992)	ACTEQ5
Training	• To indicate if they had completed any computer-related course (Dambrot et al., 1985) • Whether they have ever done a course requiring the use of a computer (Jones and Clarke, 1995) • The number of computer-related classes taken (Szajna and Mackay, 1995)	ACTEQ6

Ownership questions should be used with caution as young participants may misunderstand the concept of owning an object. The question 'do you own a computer?' may be answered yes, if there is a computer at the child's home, even if they are not allowed to use it.

Having more access to computers will make it easier for a person to increase their frequency of use. In a survey carried out with school teachers, Rosen and Weil (1995) found that more primary school teachers used computers with their students than teachers in secondary schools. They attributed this trend to the fact that more primary school teachers at that time had computers in their classroom than those in secondary schools.

Having formal training in dealing with computers may increase the person's skill and other elements of CE (Rozell and Gardner, 1999). Rosen et al. (1987)

reported that students who have taken a course involving the use of a computer (for non-programming purposes) had lower computer anxiety than those that did not. Similarly, Jordan and Stroup (1982) found a decrease in computer anxiety in students after taking part in an introductory course. In contrast, some studies have shown that training increases computer anxiety (Siann et al., 1990; Nelson et al., 1991).

7.2.1.3 OCE 3 - Diversity of Experience

Diversity of experience is the person's usage of a variety of computing software packages (Igbaria and Chakrabarti, 1990; Jones and Clarke, 1995). This also includes tasks such as use of computer-assisted learning and familiarity with computer languages. Therefore, the diversity of experience reflects the person's level of computer expertise and, indirectly, reflects the training in computers the person has received (Smith et al., 1999). Diversity of experience questions can be divided into computer activities and computer software as shown in Table 23:

Table 23: Subgroups of Diversity of Experience and Their Example Questions

Subgroups of diversity of experience	Examples	Question Code (See Section 7.3)
Computer activity	• Indicate to what extent they had worked with word processing, spreadsheet programs, programming or computer games before attending college (Busch, 1995) • Prior computer/technology experience in eight areas (automatic banking, word processing, as a student, learning programming, on the job, in a library, playing video arcade games, and playing computer games) (Rosen and Weil, 1995)	ACTEQ7
Computer software	• The variety of types of computers used (Todman and Lawrenson, 1992) • The extent to which the subject has used certain types of computer programs (Szajna and Mackay, 1995)	ACTEQ8

Diversity of experience is often asked with a list of items to be selected by the participants, or a list of specific questions covering many activities/software. An important distinction should be made here between depth and diversity of experience. A person could have an in-depth understanding of one particular piece of software, or a relatively shallow understanding of how to use several pieces of software. The CE of these two people are clearly different. In this instance, it is the range of experience a person has that is measured, rather than the depth at which they are familiar with each task/software.

Szajna and Mackay (1995) made a clear distinction between computer activities and computer software and in their survey these were presented in separate questions. Table 24 shows the items covered by their questions.

Table 24: List of Computer Activity and Computer Software Items as Defined by Szajna and Mackay (1995)

Computer Activities	Computer Software
1. Played games 2. Used packaged programs 3. Wrote programs 4. Operated a mainframe 5. Repaired computer or video games 6. Sold computers or software 7. Designed computer hardware 8. Managed computer personnel	1. Statistical package 2. Word processing 3. Graphics 4. Music 5. Accounting, financial 6. Engineering, architectural 7. Medical 8. Other

Although these lists of computer software and activities cover a wide range of computer use, it is lacking in the latest computer trends. For example, popular social networking and photo-sharing websites such as Facebook and Flickr do not appear on the list. Lists that relate to fast moving technologies should be kept up to date.

7.2.1.4 OCE4 - Sources of Information

Sources of information refers to the sources through which access to computer-relevant information has been acquired, including the media, peers, parents and teachers (Jones and Clarke, 1995). These sources could be direct, e.g. media, peers, parents, teachers (Igbaria et al., 1995) or indirect, e.g. observing, reading or hearing about another person's computing experience (Eagly and Chaiken, 1993) as shown in Table 25:

Table 25: Example Questions of Sources of Information

Aspect	Examples	Question Code (See Section 7.3)
Sources of information	• I talk about computer games with my friends (Colley et al., 1994) • The media question asked if seven media types (radio, films, newspapers, magazines, books, museums, and advertisements) had influenced the subject's current feelings about technology (using a rating scale) (Weil et al., 1990)	ACTEQ9

Questions regarding sources of information usually list the possible sources, and either ask questions for each source, or ask the participant to select all the sources of information that apply to them.

7.2.2 Subjective Computer Experience (SCE)

Subjective computer experience (SCE) is a private psychological state regarding the person's thoughts and feelings about some previous or existing computing event (Eagly and Chaiken, 1993). A person's SCE can be based on actual interaction with a real computer (direct SCE) or entirely based on what the person reads, has seen others use, or discusses with others about computers (indirect SCE) (Eagly and Chaiken, 1993). Since a private psychological state cannot be directly measured, this measure of CE is reliant upon the person's ability to accurately self-report their thoughts and feelings.

7.2.2.1 SCE1 - Perceived Competency

Perceived competency is the personal judgment of one's ability for performing a specific task (Murphy et al., 1988; Schunk, 1989). In other words, how good someone thinks they are at doing something. It is also referred to as self-efficacy, and is thought to directly impact the choice to engage in a task, the effort that will be expended and the persistence that will be exhibited (Bandura, 1977; Schunk, 1985). Bandura suggested that individuals must feel confident in using the technology to effectively employ it (Bandura, 1977). In addition, high correlations have been found in studies between self-efficacy and subsequent cognitive performance (Bandura and Adams, 1977; Bandura et al., 1977).

Questions related to perceived competency can refer to the participant's view of their skill levels, either in the general computer use, in use of a particular software package or specified tasks. Table 26 shows the subgroups and examples of perceived competency.

Table 26: Subgroups of Perceived Competency and Their Example Questions

Subgroup of perceived competency	Examples	Question Code (See Section 7.3)
In general	• I am confident about my ability to do well in a task that requires me to use computer technologies (Kinzie et al., 1994) • Computer confidence - 'very low' to 'very high' (Weil et al.,	ACTEQ10

	1990)	
Particular software	• How students would rate their skill in using the computer for word processing (Arch and Cummins, 1989) • The acquired level of skill in using these applications self-rated by the participants (Beckers and Schmidt, 2003)	ACTEQ11
Particular task	• I feel confident moving blocks of text while word processing, etc. (Ertmer et al., 1994) • I feel confident logging on to e-mail, etc. (Kinzie et al., 1994)	ACTEQ12 ACTEQ13 ACTEQ14 CCTEQ4 CCTEQ5

Self-efficacy is considered to be situational-specific (Murphy et al., 1988). It is possible for someone to perceive himself or herself as being not very good at using computers, but very good at typing. It is thus important that care is taken to ensure these questions are directed at particular items of interest. Questionnaires are available that focus solely on gathering self-efficacy data, such as the *Self-efficacy for Computer Technology* questionnaire (Delcourt and Kinzie, 1990).

7.2.2.2 SCE2 - Control

The aspect of control mainly covers how in control a person feels or felt when using a computer. The question of control can be aimed at how in control they feel currently when using a computer (Kinzie et al., 1994), or how in control they felt during early experience with computers (Todman and Monaghan, 1994) as shown in Table 27:

Table 27: Example Questions on Control

Aspect	Examples	Question Code (See Section 7.3)
Control	• Whether, during early experience with computers, the participants had (a) generally felt pretty much in control, (b) generally felt more or less in control much of the time or (c), generally felt that he or she was seldom in control (Todman and Monaghan, 1994) • The extent to which early experience was 'hands on' (Todman and Monaghan, 1994)	ACTEQ15 ACTEQ16

Todman and Monaghan (1994) found that control during early experience had significant correlation with current frequency of computer use. They also found that the more in control the participant felt in early experience, the less anxious they felt and was more likely to continue using the computer subsequently.

7.2.2.3 SCE3 - Perceived Usefulness

Perceived usefulness (Table 28) refers to the person's subjective opinion on whether a particular system/technology will increase their job performance (Davis et al., 1989). They suggested that perceived usefulness strongly influenced people's motivation to adopt a new system and thus gain more CE.

Table 28: Example Questions on Perceived Usefulness

Aspect	Examples	Question Code (See Section 7.3)
Perceived usefulness	• I don't have any use for computer technologies on a day-to-day basis (Kinzie et al., 1994) • With the use of computer technologies, I can create materials to enhance my performance on the job (Kinzie et al., 1994) • If I can use word-processing software, I will be more productive (Kinzie et al., 1994)	ACTEQ17 ACTEQ18

Both Igbaria et al. (1995) and Kinzie et al. (1994) found self-efficacy for computer technologies to be positively related to perceived usefulness and frequency of use, suggesting that when a person perceives something to be useful, they are likely to use it more, and become more proficient at it.

7.2.2.4 SCE4 - Attitude

Although not included in Smith et al.'s (1999) categories of SCE, attitude towards computers is frequently reported as having a strong positive correlation with the person's experience in computers (Kinzie et al., 1994). Many questions asking the participant's attitude towards computers have been found in literature (Weil et al., 1990; Kirkman, 1993; Ertmer et al., 1994; Kinzie et al., 1994) and there are several questionnaires that focus on gathering computer attitude, such as the *Attitude Towards Computer Technologies* questionnaire (Delcourt and Kinzie, 1990), and *Computer Anxiety Scale* questionnaire (Newman and Clure, 1984), some of the questions from which are listed in Table 29.

Table 29: Example Questions on Computer Attitude

Aspect	Examples	Question Code (See Section 7.3)
Computer attitude	• How much students liked computers (Kirkman, 1993) • How much they enjoyed using computers (Kirkman, 1993)	ACTEQ19 CCTEQ6

	• I feel at ease learning about computer technologies (Ertmer et al., 1994) • I am anxious about computers because I don't know what to do if something goes wrong (Kinzie et al., 1994)	

In general, computer attitude increases as the person's CE increases (Hill et al., 1987), and together these positively affect their self-efficacy (Delcourt and Kinzie, 1993). Loyd and Fressard (1984) identified three types of computer attitude that seem to have the most significant effect on achievement of computer tasks by students: anxiety, liking and confidence. It is usually assumed that those with more CE have less anxiety about using computers and thus feel more in control. There are many studies that support this assumption (Okebukola et al., 1992), but also numerous studies exist that find no relation between the two factors (Kay, 1990; Todman and Lawrenson, 1992) or even negative relationships (Howard and Smith, 1986; Heinssen et al., 1987; Igbaria and Chakrabarti, 1990). As the interaction between computer anxiety and CE is beyond the scope of this study, readers are directed to Kinzie et al. (1994) for a review of literature in this topic.

7.2.3 Shortfalls of Current Questionnaires

This section showed that there are many methods designed to gather a participant's CE. However, most focused in depth on measuring a few aspects of CE, and failed to provide a general overview. There are many aspects of CE that could be asked to the participants. However, sometimes it is not possible to ask so many questions regarding just one thing. The CE questions may only be a small part of the study, such as in this thesis. It would be far more useful to know which aspects most correspond to typing performance, so that in future typing studies a select few questions that are most indicative could be used.

Additionally, some of the questions, particularly those regarding what tasks they have performed before, or what software they have used, were found to be outdated. These questions require updating to modern tasks, such as the use of Facebook and YouTube.

Finally, very few questions were directly related to the participant's typing. In a typing study, it would be far more interesting to see if any of them had formal training in typing, or how they felt about their typing ability. It is possible for someone to have a negative attitude towards computers but be highly skilled at typing. A questionnaire used for this study should contain questions directed to typing itself.

7.3 DESIGNING THE ADULT QUESTIONNAIRE

To address some of the shortfalls in existing CE questionnaires, a new questionnaire, the Computer and Typing Experience Questionnaire (CTEQ), was designed. The Adult CTEQ (ACTEQ) consists of 19 questions. These 19 questions were derived from the 16 subgroups of CE identified in Section 7.2, with some subgroups providing more than one question.

7.3.1.1 Objective Computer Experience (OCE) Questions

There are nine questions in ACTEQ that asked the objective aspects of CE. OCE are grouped by a*mount of computer use, opportunity to use computers, diversity of experience,* and *sources of information.*

For the amount of computer use, there were three questions:

- ACTEQ1: *How old were you when you first used a computer?*
- ACTEQ2: *How often do you use a computer?*
- ACTEQ3: *On the days you do use a computer, how many hours do you use it for?*

These three relate to the amount of computer use (Section 7.2.1.1). ACTEQ1 relates to the duration of computer use in the participants' lives. ACTEQ2 relates to the frequency of computer use and the answer is given on a 5-point scale of 'every day', 'several times a week', 'once a week', 'several times a month' and 'less than once a month'. ACTEQ3 asked about the duration of computer use per day, and the participants had to choose from a 5-point scale of 'all day', 'several hours', 'one hour', 'half an hour' and 'less than half an hour'.

Next, three questions related to the three subgroups of opportunity to use computers (Section 7.2.1.2):

- ACTEQ4: *How many laptops/computers do you own?*
- ACTEQ5: *Do you have access to computers at your work/university/college/school?*

- ACTEQ6: *Have you ever had formal lessons in how to type?*

ACTEQ4 is a question related to the ownership of computers where the participants were expected to enter the number owned. ACTEQ5 is a yes or no question regarding the access of computers. ACTEQ6 asks about prior training. In previous questionnaires regarding a person's general CE, this last question usually asked about any training the participant have received in using the computer or for particular software. However, with IT lessons being a core part of the National Curriculum for many years, a question asking whether or not the participant had formal training in the use of a computer seemed obsolete. In contrast, fewer and fewer people are receiving formal touch-typing lessons. Therefore, this question was changed to training specific to typing. The answers available for this question were a yes or a no.

Two questions asked about the participants' diversity of experience (Section 7.2.1.3). For the computer activity task question (ACTEQ7), the TypingCollector asked the participant to select all tasks they have done before on a computer. There were ten tasks in total, which range in level of expertise required, from common computer user tasks (such as playing a computer game) to specialised tasks that were more commonly carried out by developers:

- Played computer games
- Edited photos
- Uploaded a video or photo to a website
- Made a presentation
- Created a spreadsheet
- Created a spreadsheet using formulae
- Developed a website
- Created a blog
- Built a Flash application
- Written a computer program (C++, Java, etc.)

If a participant selects 'created a spreadsheet using formulae', then they must also select 'create a spreadsheet'. If they do not, then it is possible that the participant's data contain random selection errors.

For the computer software question (ACTEQ8), the TypingCollector asked the participant to select all software they have used before. Traditionally, these only included software, but with the wide spread use of websites such as Facebook and Flickr, these were added to the list of 'software'. Like the previous question, the options given to the user cover a wide range of computing expertise.

- Social networking site (e.g. Facebook)
- Image sharing websites (e.g. Flickr)
- Blog builder (e.g. Blogspot)
- Word processing software
- Spreadsheet
- Webmail
- Image manipulation software (e.g. Photoshop)
- Flash builder
- Integrated development environment (IDE)
- Statistical analysis software (e.g. SPSS)

This question also contained some checking mechanisms to test whether or not the participant was selecting answers randomly. It stood to reason that if someone selected 'created a spreadsheet' and/or 'created a spreadsheet using formulae' in ACTEQ7, then they must also select 'spreadsheet' in ACTEQ8. Similarly, if they have selected 'built a Flash application', they must also select 'Flash builder' in ACTEQ8. If the participant selected 'edited photos' in ACTEQ7, they should also select 'image manipulation software (e.g. Photoshop)' in ACTEQ8.

The Final OCE question was the sources of information (Section 7.2.1.4). This question asked:

- ACTEQ9: *What resources do you use when you are learning something new on a computer?*

The participants could select any number of the following options: friends, family, at work, at school/college/university, websites, blogs, online forums, books, magazines or instruction manuals. Blogs and online forums were added to the traditional range of sources, since these sources are becoming more widely used.

7.3.1.2 Subjective Computer Experience (SCE) Questions

There were four main aspects of SCE - *perceived competency, control, perceived usefulness* and *computer attitude*. These were broken down into ten questions.

There were three subgroups in perceived competency (Section 7.2.2.1) - in general, in a particular software and in a particular task. Since this questionnaire places its focus on typing, the questions were orientated towards typing tasks.

- ACTEQ10: *How good do you think you are at using computers?*
- ACTEQ11: *How good do you think you are at using word processing software (e.g. Microsoft Word)?*
- ACTEQ12: *How good do you think you are at typing?*
- ACTEQ13: *How fast is your typing?*
- ACTEQ14: *How accurate is your typing?*

The last subgroup, perceived competency in a particular task, were split into three questions. This was to see if there were variations between their own perceptions of how good, how fast, and how accurate they are at typing. All four questions used a 5-point scale, with ACTEQ10, 11, 12 and 14 using 'very good', 'good', 'ok', 'not very good' and 'poor'. ACTEQ13 used the options 'very fast', 'fast', 'average', 'slow' and 'very slow'.

The subject of control (Section 7.2.2.2) was separated into two questions, one on how much in control they feel when they are using it, and the other on how much in control they felt they were regarding their learning of how to use a computer:

- ACTEQ15: *How in control do you feel when using a computer?*
- ACTEQ16: *How much control did you feel you had over how you learnt to use computers?*

Both questions used a 5-point scale with 'very much in control', 'in control', 'OK', 'not in control' and 'totally out of my control' as their options.

Perceived usefulness of the computer (Section 7.2.2.3) was separated into two contexts, in terms of usefulness to their work or education, and usefulness to their everyday life. It was possible for a person to view a computer as being absolutely crucial for their work, but perhaps not for their everyday life and vice versa.

- ACTEQ17: *How useful do you think computers are to your work/education?*
- ACTEQ18: *How useful do you think computers are to your everyday life?*

Both questions used a 5-point scale with 'very useful', 'useful', 'neither useful nor useless', 'useless' and 'very useless'.

Finally, a question regarding their attitude towards computers was asked:

- ACTEQ19: *How much do you like using computers?*

The question used a 5-point scale of 'really enjoy', 'like', 'OK', 'dislike' and 'hate it'.

7.4 Study 1 - Evaluating the Adult CTEQ

For the purpose of this thesis, computer experience is used as a tool to investigate which of the CE aspects are affected by demographics of the sample selected. This study investigates a large sample of the student participants to establish which particular subgroups of CE were affected by the fact that they were first-year undergraduate computing students, and which aspects remained normally distributed.

7.4.1 Method

A five-day study was carried out to evaluate ACTEQ with first-year undergraduate computing students. A question from each subgroup of CE found was used to build a detailed Adult CTEQ. Answers to these questions were collected digitally from undergraduate computing students from the author's department, along with some phrase-copy typing data. Comparisons were made on correlations between answers to each question and their typing performance.

7.4.2 Participants

137 first-year undergraduate computing students took part in this study. As with most other computing degree courses, the samples were heavily skewed towards the male population with 124 males and 13 females in total. The age range was 18-43 years, but most (134) were between 18-30 years old. These participants are represented as group 15, participant ID S93-S229 in the participant summary chart found in Appendix 3.

7.4.3 Apparatus

The computer laboratories used in this study all had identical PCs and keyboards. Since the data collection was carried out in the specific rooms that the students normally used for their laboratory sessions, everyone was familiar with the equipment.

The TypingCollector introduced in Chapter 6 was adapted to ask CTEQs following the demographic questions. Digitisation of paper-based questionnaires have been widely studied in adult participants (Naus et al., 2009). They have found that there is virtually no difference in answers between paper and computer-based questionnaires (Petitt, 2002; Truman et al., 2003; Vereecken and Maes, 2006; Naus et al., 2009). Therefore, it was deemed reasonable to give a computer-based version of ACTEQ to the adult participants.

7.4.4 Procedure

The study was carried out at the start of five practical classes of the same computing first-year undergraduate course. There were approximately 30 students per lab, and each had access to a PC. At the start, the procedure was explained to the participants. They were asked to download the TypingCollector from their module resource webpage, and save it on their desktop to run it. Once the TypingCollector had ended, the participants were asked to upload their log file onto their module resource webpage.

7.4.5 Analysis

For questions that used 5-point scales, answers were scored from 5 to 1, with 5 being the most positive responses (e.g. 'very good', 'very fast', 'very useful') and 1 being the least positive responses (e.g. 'very bad', 'very slow' 'very useless').

For the diversity of experience questions (ACTEQ7 and 8) where the participants selected a number of tasks/software, a score was given for the number of selections made. Although it was possible to weigh the items according to the expertise required (making the more 'expert' tasks count for more), instead, an equal weighting was placed on each question. This assumes that if a participant selected all 10 tasks/software, then they had a higher CE than those that selected fewer tasks/software. The same method of scoring was applied to the sources of information question (ACTEQ9).

A simple CE score was calculated for participants by totalling their score for each question. Years of computer use (ACTEQ1) were removed from the total score for reasons explained in Section 7.2.1.1. Although all participants gave answers to all the ACTEQs, in the phrase-copying task, one person (S208) typed the notification of what number of phrases they are on (e.g. '1 out of 10') rather than the presented phrases. Therefore, this person was removed from the sample.

7.4.6 Results

Figure 16 below shows that the total CE score for the adult sample was normally distributed (Kolmogorov-Smirnov test, $D(136) = .059$, $p > 0.150$). However, studying the individual ACTEQs revealed that answers given to several ACTEQs were not normally distributed.

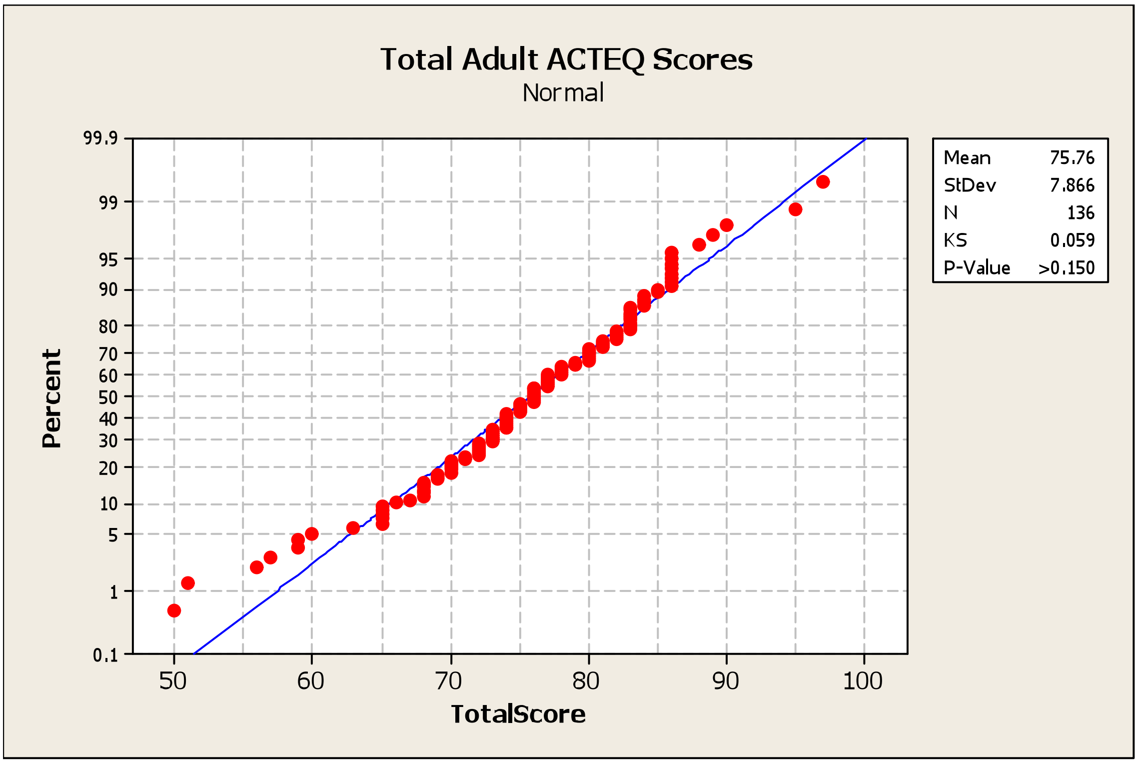


Figure 16: Normal Q-Q Plot for Total CE Score for Adults

Table 30 below shows results of Kolmogorov-Smirnov test for each ACTEQ item in the adult questionnaire. Items that were not normally distributed have been highlighted in bold. ACTEQ5 - access to computers at the workplace/university has been removed from the table since all participants answered 'yes'. ACTEQ6 - whether the participant had formal typing training or not - was also removed from the table since this was a binary yes/no question.

Table 30: Results of Kolmogorov-Smirnov Test on Individual ACTEQs (Items that were not Normally Distributed has been Highlighted in Bold)

	Item	Mean	*D*(136)	*p*
ACTEQ 2	**FrequencyUse**	**4.919**	**0.079**	**0.041**
ACTEQ 3	Duration	4.088	0.048	> 0.150
ACTEQ 4	**Ownership**	**2.301**	**0.087**	**0.017**
ACTEQ 7	**TaskScore**	**8.279**	**0.115**	**< 0.010**
ACTEQ 8	SoftwareScore	6.838	0.029	> 0.150
ACTEQ 9	InfoScore	5.228	0.026	> 0.150
ACTEQ 10	CompetenceGeneral	4.382	0.065	> 0.150
ACTEQ 11	CompetencySoftware	4.390	0.066	> 0.150
ACTEQ 12	CompetencyTask	3.956	0.031	> 0.150
ACTEQ 13	CompetenceTypingSpeed	3.588	0.043	> 0.150
ACTEQ 14	CompetenceTypingAccuracy	3.684	0.050	> 0.150
ACTEQ 15	**ControlUse**	**4.500**	**0.081**	**0.035**

ACTEQ 16	ControlLearning	3.941	0.034	> 0.150
ACTEQ 17	UsefulWork	3.941	0.034	> 0.150
ACTEQ 18	UsefulLife	3.941	0.034	> 0.150
ACTEQ 19	**Attitude**	**4.800**	**0.091**	**< 0.010**

Items that had positively skewed distribution were access to computers in the workplace/university, frequency of use, number of computers owned, task score, feeling of control when using a computer and attitude. In particular, the mean task score and attitude to computing were high.

7.4.7 Discussion

7.4.7.1 Items Affected by the Sampling Method Used

For a sample of undergraduate computing students, the question of how frequently they used a computer (ACTEQ2), and of whether they had access to computers at university (ACTEQ5) were redundant questions. Unsurprisingly, for frequency of use, most participants selected either 'several times a week' to 'everyday', and for access, all selected 'yes'. However, these questions may provide finer granularity in terms of CE in a different study where the sample is selected from a wider background.

Similarly, it was not surprising that computing students felt more in control of using a computer and had positive attitude to computing in general. The number of computers they owned also had a positively skewed distribution. Although most answered that they owned one or two computers or laptops, there were several participants that reported owning 5+ computers or laptops. The question of ownership may be refined further to ask how many computers they currently use, rather than how many they own in total, since it is possible to own ten laptops, but only be using two of them.

7.4.7.2 Task and Software Scores

The positive skew in the distribution of the task score (ACTEQ7) was expected from computing students. However, it was also expected for the students to have a positively skewed distribution on the software score (ACTEQ8) as well. The difference in the distributions of the two items suggests that they did not cover the same range of previous experiences. Indeed, the software question contains items such as the statistical analysis software, for which the task you would carry out using it is not listed in the task question.

Further, it is possible that the participants did not know the formal names of the software they have previously used. For example, it is entirely possible that a student using Netbeans to write their Java program may not know that this software is referred to as IDE (Integrated Development Environment) in general. In such case, the student would select 'written a computer program' in the task question, but not select IDE in the software question.

It is suggested that a further work should be carried out to make the task and software questions to encompass the same range of computer experience. Furthermore, to ensure more accurate answers, names used to refer to each software should be tested to ensure that the participant answering the questionnaire understands each of them.

7.4.8 Conclusions

By examining each of the 19 aspects of CE, it has been possible to understand where the sampling method has been affected by the sample's previous computer experience. The ACTEQ allowed for narrowing down the aspects that were affected by the demographic, and which aspects maintained a normal distribution, providing a much more detailed understanding of the sample.

However, care should be taken in generalising these findings implying that the sample was representative of the population. Even if the sample used in this study had shown normal distribution in all aspects tested in the adult questionnaire, it would not have meant that they were representative of the general population. It can only be inferred that the sample had a normal CE distribution within itself.

7.4.8.1 Limitations

Clearly, the participants all being computing students is a limitation in this study. It is likely that people may perceive themselves as less competent at using a computer if highly competent users surround them. In contrast, it could be argued that someone would regard himself or herself as being good with computers if they have enrolled on a computing degree.

As much as possible, the questions were designed to reduce this by asking participants actual figures (such as number of days they use a computer for frequency of use) rather than subjective scoring (such as 'never', 'often', 'very often'). However, it is important that a much larger study, with a wider range of participants in terms of computer use should be carried out before any generalisations are made on what ACTEQ items truly correlate most with.

Although adults have less problems understanding the concept of ownership than children do, the ACTEQ4 - 'how many laptops/computers do you own?' was vague. It was possible for the participants to interpret this question in many ways - such as counting all computers in the household, or ones that they no longer use but are still in their household. This question should be more clearly defined, such as asking for how many computers and laptops they regularly use.

7.4.8.2 Contributions

The contribution of this study is the construction and evaluation of an up-to-date and in-depth questionnaire that asks participants about their CE and typing. The questionnaire allowed for highlighting those aspects that were biased by the demographic of the sample. The separation of those CE aspects that had either a negatively or positively skewed distribution from those that were normally distributed allows for a better understanding of where the bias in the sample comes from.

7.5 Study 2 - Pilot Study for the Child CTEQs

Whilst several methods for measuring CE exist for use with adults, these methods are seldom adapted from their original form when used with children.

Questionnaires have been used with young children in Child Computer Interaction with success (Scott, 2000; Markopoulos et al., 2008). They are popular as they can be administered to large numbers of children (e.g. a whole class) simultaneously with relatively low workload for the investigators (Markopoulos et al., 2008).

One disadvantage of using questionnaires, as opposed to other survey techniques such as interviews, is that it is not usually possible to ask the participant to clarify their answers (Scott, 2000). Markopoulos et al. (2008) highlight that difficulties often encountered in using questionnaires with children include ensuring that the children understood the question asked and in eliciting accurate answers from children.

Before asking young children about any question, it is crucial that children understand the topic in discussion (Read et al., 2001). Borgers et al. (2004) also recommend the use of children's 'own words' when creating questionnaires for children. They also warn of issues that can occur when using negatively phrased questions or including any level of ambiguity.

Therefore, researchers must first ensure that the children understand the concept presented to them in a questionnaire. For example, when asking how often a child uses a word processor, the researcher must first know that the child knows what is meant by 'a word processor'. To do this, in a questionnaire on school delinquency, Loeber and Farrington (1989) first asked their young respondents for examples of each concept (e.g. skipping school), and only used the answer if the example given was correct.

Besides understanding the questions, the children may also have difficulties understanding how they should respond. Scott (2000) suggests that for children under 11, the use of visual stimuli is useful in making a concept in question more concrete than verbal representation alone. Read et al. (2001) applied this to scales used to measure various concepts with young children (6-10 years old). In particular, the Funometer and the Smileyometer used a Visual Analogue Scale (VAS) to assist children in understanding what the scale represented.

7.5.1 New Scales

Two new scales, one VAS and one rating scale, were developed for use in the Children's CTEQ. These scales have been designed to assist children in answering questions regarding their CE more accurately.

7.5.1.1 Thumbs-Up Scale (TUS)

The Thumbs-Up Scale is a VAS designed to measure the children's perception of their skill in a particular task, as in the example case described below – the skill being measured was typing.

Figure 17 shows an example TUS asking the participant to indicate his or her perceived typing skill. The TUS may be applied to any question regarding how the children perceive themselves as being good or bad at a particular skill.

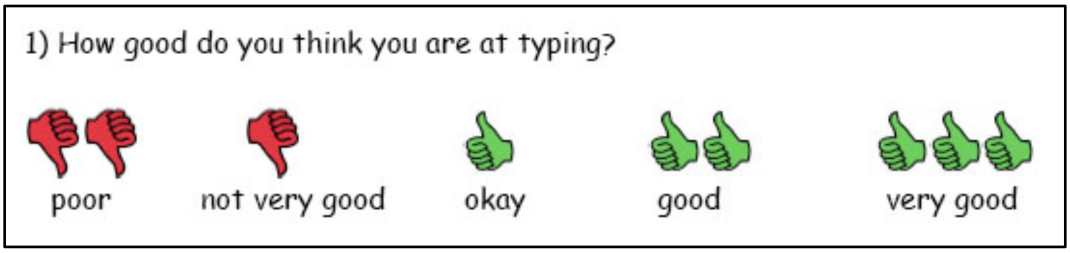


Figure 17: An Example Use of the Thumbs-Up Scale (TUS)

7.5.1.2 Frequency of Use Scale (FUS)

Since frequency of use is one of the major factors in a person's CE, it was important to ensure that the scale measuring frequency was suitable. For this, a 4-point scale was designed, with 'never', 'once a week', 'a few times a week' and 'every day'. This scale may be adapted to any task that children carry out. Although the example here is for a task that occurs at least on a weekly basis, these measures can be altered for monthly or annual tasks.

7.5.2 Method

A one-day study involving 49 children from two local primary schools was carried out to investigate whether or not young children understand questions relating to computer hardware and software, and whether the use of TUS and FUS are appropriate for children aged seven and upwards. Carrying out pilot studies for questionnaires for children is important in ensuring that the questions are suitable for them (Markopoulos et al., 2008). TUS and FUS were worded to ask questions relating to typing as a test case for these scales.

Children in the study completed a paper-based questionnaire consisting of 12 questions, followed by carrying out the 10-phrase-copying task on the Data Collector.

7.5.3 Participants

There were 24 boys and 25 girls. 26 were from Year 3 (7 to 8 years old) and 23 were from Year 5 (9 to 10 years old). Both classes were chosen from two local primary schools in Lancashire, UK. The Year 3 class is identified as group 9 in the participant summary found in Appendix 3, participants are identified as C117 to C140. The Year 5 class is identified as group 10, with participant IDs between C141 to C163.

7.5.4 Procedure

The two classes visited the university separately but on the same day – each class was from a different school. The children were selected by their teachers to be in groups of three. Several activities were available in separate rooms and the children took turns to carry out the tasks. The questionnaire activity was carried out in a quiet computer lab with five to seven children at a time. Each child was asked to fill in the questionnaire individually with a pen. They were then assigned a computer each to sit in front of. The TypingCollector then ran a 10-phrase-copying task for each participant.

7.5.5 Design

The paper-based questionnaire had two questions relating to CE, across two pages, written in Comic Sans font size 12pt. The questions consisted of seven core questions related to the participant's CE, three questions checked their understanding of the concepts in question, and two questions were designed to validate TUS and FUS.

The seven core questions asked about the children's CE. The first of these asked the child's own opinion on his or her typing skill (SCE) and used TUS. The second question asked how much they liked typing (SCE) and used a Smileyometer (Read et al., 2001) as shown in Figure 18. Three questions then asked about the frequency of computer use at home, at school and use of word processor (OCE). All three questions used the same FUS scale ranging from 'never' to 'everyday'. The questionnaire also asked whether or not the children had a computer or a laptop in their own bedrooms and asked how old they were when they first used a keyboard (both relating to OCE).

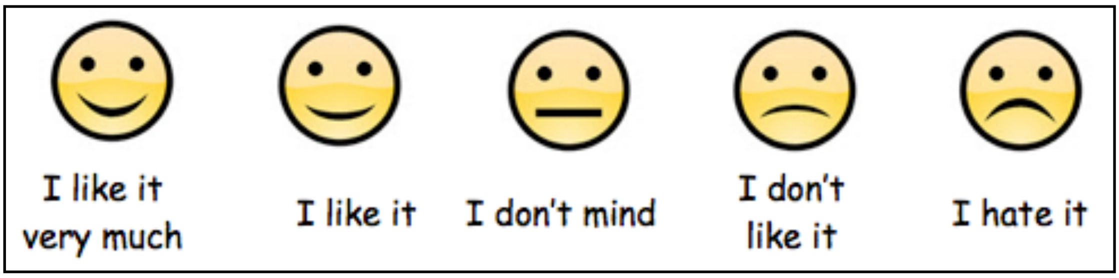


Figure 18: An Example of a Smileyometer Used in the Children's Questionnaire

To ensure that children understood the concepts in question, three questions were designed to see whether or not the children had an appropriate understanding of relevant computer software and hardware. The first of these showed pictures of a computer mouse, a PC, a laptop and a keyboard and asked the children to name these devices. The other 2 questions asked the child to write down what tasks could be done with a computer keyboard and a word processor.

The two final questions were designed to validate the answers the children gave using TUS and FUS. These essentially asked the same question, but the options for answering were presented differently using a cloud of words from which the child was required to select an answer. It was assumed that if the children gave similar answers on both questions, then the new scales were appropriate for children. The new scales were presented at the start whilst the validation questions were positioned at the end of the questionnaire.

7.5.6 Results

All the children were able to complete the question about their skill (the TUS) with the results being 65.3% ‘good’ or ‘very good’, 22.0% 'okay', and 12.3% 'not very good' or 'poor'. The Smileyometer was easily completed by all children with 77.6% responding 'I like typing' or 'I like it very much' and only 8.2% reporting 'I don't like it' or 'I hate it'. The other questions in this part also had a 100% completion with all the children reporting using a computer at school, with older children using it more frequently. Only two reported using it everyday in school. 86% of children reported using a computer a few times a week to everyday. Only two children reported never using a computer at home.

7.5.6.1 Children’s Understanding of Computer Hardware and Software

From the first question in part two of the questionnaire, 48 out of 49 children were able to name all four devices correctly. This indicates that children as young as seven have a good basic understanding of computer hardware, and so it is appropriate to ask questions regarding these. In question two of this section, 47 children answered the question 'what do you do with a computer keyboard' correctly with answers such as 'type' and 'type words.'

However, in the third question in this part, children indicated some lack of understanding of computer software. Eight children could not answer the question 'what can you do with a word processor (Microsoft Word)'. Many children indicated difficulties with this question during the study, asking the researcher what a word processor was. The children were more familiar with the software name of 'Word', rather than 'word processor' and generally comprehended the question fully when it was restated as 'what can you do on Word?' It is suggested that ‘word processor’ should not be used by young children but refer in the question to names of products they are familiar with.

7.5.6.2 Scale Validations

The TUS and its validation word cloud scale, both measuring the children’s perceived skill level, had a high correlation (r=0.892). Despite the two scales having exactly the same range of answers, children scored themselves lower on the word cloud scale (M=0.67) than in the thumbs-up scale (M=0.76).

To validate FUS, the question of frequency of computer use at home (previously asked with a FUS) was asked again but using a cloud diagram with numbers from 0-8 (representing the number of days in a week) scattered randomly in a box. The number 8 was added as an option to see if children would use 8 to mean that they use the computer very frequently. For this comparison, the numerical values of 0 to 8 were recoded so that they meant ‘never’ = 0, ‘once a week’ = 1 and 2, ‘a few times a week’ = 3, 4 and 5, and ‘everyday’ as 6, 7 and 8. The two scales had a high correlation (r=0.744). Overall, children rated themselves similarly on both scales.

These high correlations indicate that the new scales measured their respective items in very similar manner to the cloud diagrams. Children are able to answer questions using these scales with similar accuracy to if they were using cloud diagrams.

7.5.7 Conclusions

Two new scales were proposed to assist children in measuring their own CE more accurately. These scales were tested for validity within a pilot questionnaire designed to quantify CE.

It was found that children as young as seven years old were able to understand and effectively use the TUS for perceived skill levels and FUS for frequency, indicated by the high correlations between the new scales and their respective validation measures.

Additionally children understood computer hardware enough to be able to answer questions relating to the basic concepts of computers. However, the children had a weak understanding of software, suggesting that CE questions regarding tasks and software usage may not be suitable for young children.

7.5.7.1 Limitations

All the children that participated in this study were recruited from two local primary schools. Additionally, there were relatively few participants (49), which is large enough for the findings to be useful, but cannot be generalised to all children.

7.5.7.2 Contribution

The study has revealed that young children are able to answer questions regarding their computer usage, and perceived competencies. However, they struggled with questions that referred to named tasks, suggesting that questions regarding the range of tasks and software usage are not suitable for young children.

7.6 DESIGNING THE CHILD QUESTIONNAIRE

The Child Computer and Typing Experience Questionnaire (CCTEQ) contain 6 questions to collect previous computer experiences from children. The questionnaire contains considerably fewer questions than in the adult's one (adults had 19 questions).

The questions in the CCTEQ were:

- CCTEQ1: *How old were you when you first used a computer keyboard?*

This was an OCE question relating to the OCE measurement for amount of computer use (Section 7.2.1.1). The children entered the age at which they first used a computer keyboard. The previous study (Section 7.5) had shown that children a year younger than this participant group understood what was meant by a computer keyboard. The computer-based version did not allow input greater than the participant's given age.

- CCTEQ2: *Do you have a computer or a laptop in your bedroom?*

This was another OCE question, this time relating to the opportunity of use (ownership - Section 7.2.1.2). In the adult questionnaire, this question was simply 'how many computers and laptops do you own?' However, children often have difficulty in understanding ownerships of properties. For example, a child may have a computer in the living room in their house that they think is their own, but perhaps it is used mostly by the parents for their work. Thus it was decided that if a child has a computer or a laptop in their own bedroom, it was deemed that it belongs to them. This question was set out as a yes/no question.

- CCTEQ3: *How many days a week do you use a computer at home?*

This was the third OCE question asked regarding the amount of computer use. Since schools have similar number of hours set aside for ICT classes (following the key stage guidelines), the amount of use at home is likely to provide a wider variety of answers. The answer to this question was set out as a 5-point scale of 0 days/1 to 2 days/3 to 4 days/5 to 6 days/7 days.

- CCTEQ4: *How good do you think you are at typing?*

This question relates to the SCE regarding their perceived competency (Section 7.2.2.1). This question used the TUS scale tested in the previous study in Section 7.5 (see Figure 17).

- CCTEQ5: *How fast can you type?*

This was another SCE question regarding their perceived competency, and used the TUS scale again. This question differs slightly from the previous question and was felt important to ask, for cases where a child may feel that they are good but careful (slower) typists. The same Thumb-Up images were used as the rating scale, but the words were changed to 'very fast', 'fast', 'average', 'slow' and 'very slow'.

- *CCTEQ6: How much do you like typing?*

This was an SCE question regarding computer attitude (Section 7.2.2.4) and used the Smileyometer (Read et al., 2001) as shown in Figure 19 below. It was decided to focus the question on the typing task alone, since the number of questions was limited.

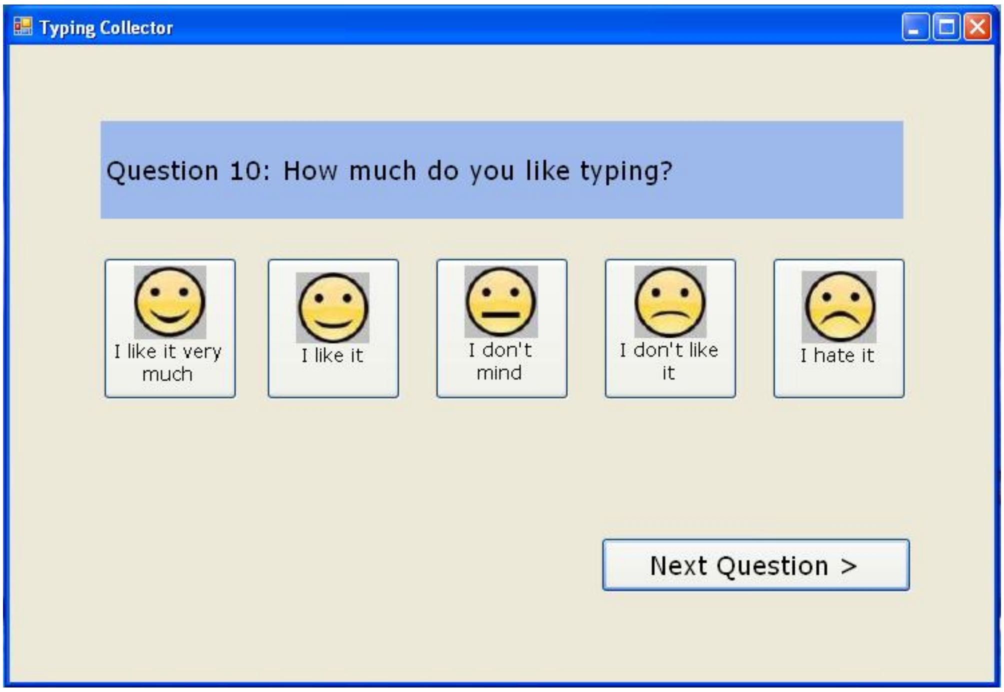


Figure 19: Smileyometer Scale in the TypingCollector

The previous study in Section 7.5 found that the children had difficulty in understanding software, so it was assumed that young children would have difficulty answering questions regarding what tasks they have carried out on the computer. It was decided that diversity of experience (OCE) questions would not be asked.

The questionnaire does not give the participants the option to say 'I don't know'. Although it desirable to have the option of 'I don't know' if it is the true state of that person's answer to the question, this noncommittal option discourages children from expressing their opinions by offering an easy way out (Bell, 2007).

7.7 Study 3 - Comparison of Paper-Based and Computer-Based Questionnaire with Young Children

A computer-based questionnaire was desirable for this work for two reasons. One was to reduce the time and effort required to co-ordinate the gathering and combining of questionnaire answers with typing data. Often, in text input studies, the questionnaires are paper-based and filled in separately to the typing data. Investigators must therefore ensure that a mechanism is in place to keep the digital data and paper-data in matching order.

The second reason was that children sometimes produce illegible answers in paper-based questionnaires. A computer-based questionnaire eliminates this response error since the participants enter values via the keyboard. Additionally computer-based questionnaire can be set to require the participant to enter an answer before going on to the next question and place checks to ensure that the values entered lie within a certain range.

There have been many studies carried out on the effects of administering a paper-based questionnaire and a digital questionnaire in the fields of medical science and psychology. Internet-based questionnaires are viewed as a cheaper alternative to paper, that can be completed by more people, in more places.

Petitt (2002) carried out a comparison study between paper and internet-based questionnaires regarding personality scales. She found that there was no significant difference between the two modes of administration. This finding is consistent with many other studies on other psychological and medical questionnaires such as Burke et al. (1995), Pouwer et al. (1998), and Lukin et al. (1985) to list a few. She also found the paper-based questionnaire to suffer from a statistically higher number of errors.

Naus et al. (2009) tested three questionnaires that cover distinctly different but commonly assessed areas in psychological research. They performed a within-subject counterbalanced-ordered study of 76 undergraduate female students, who took all three tests in two different formats, on paper, and on the Internet. They found no difference between the formats for two of the tests, which is consistent with seven previous studies for the two tests. However, some of the subsections of the third test showed significant differences between the formats, which disagreed from previous findings (Rammstedt et al., 2004). This inconsistency with previous data may have risen from Naus conducting four tests all together, whereas Rammstedt carried out only one test.

There are fewer studies carried out with younger participants. A Flemish study carried out in 2006 (Vereecken and Maes, 2006) tested paper and computer-based versions of a health and lifestyle questionnaire. The study was between-subjects with 1608 participants aged between 12 and 20. They found no significant differences between the two formats in the majority of questions, but found that answers to several questions regarding feelings were significantly different. The participants gave more sociably desirable answers in the paper-based questionnaire than the computer-based questionnaire. This was possibly down to the set up of the study, which were ran in a classroom, and thus other students could potentially see what was being selected.

Truman et al. (2003) tested 214 children aged between 8 and 15 years old. He found no significant differences between paper and computer based questionnaires regarding children's psychopathologically and psychological strengths. This finding is consistent with other mode of administration studies in adolescent such as Hallfors et al. (2000), Milllstein (1987) and Webb et al. (1999).

The aim of this study was to establish whether or not paper-based and computer-based modes of administering questionnaires about children's CE were interchangeable. Since mounting paper-based questionnaires to a digital format can change the structure of the measure, thus affecting the reliability of the data collected (Cronbach, 1990; Buchanan et al., 2005), an equivalence study was needed (AERA and NCM, 1999).

7.7.1 Method

A one-day study involving 20 children from a local primary investigated whether or not the different mode (computer or paper-based) of administration of CE questionnaire affects the answers given by young children. The design of this study was a within-subjects single factor study with two conditions, paper-then-computer or computer-then-paper.

7.7.2 Participants

There were 8 girls and 12 boys from Year 4. They were all aged 8 and 9 years old with one exception of a seven year old. They are identified as group 11 (participants IDs C164-C183) in the participant summary chart found in Appendix 3.

7.7.3 Apparatus

For the computer-based questionnaire, the TypingCollector (Section 6.5) was used, adapted with new questions regarding the children's CE. The TypingCollector first asked the demographic questions (age, gender, school year and handedness). Depending on the age of the participant given, the TypingCollector selects between asking the Adult CTEQ and Child CTEQ. When the participant entered an age of below 13 years (this can be changed in the software) the TypingCollector asked six Child Computer and Typing Experience Questions (CTEQ). Finally, the TypingCollector showed 10 phrases for the children to type, as before. The TypingCollector was run on identical PCs with identical keyboards (white writing on black keys) and monitors. All text shown by the TypingCollector was displayed in Verdana (a Sans-Serif font style) and font size 14pt.

The paper questionnaire consisted of exactly the same six Child CTEQs. The order of the questions, the wording, the scales and images used all remained the same for consistency. There were only two differences between the computer-based and paper-based questionnaires. One was that the computer-based questions were displayed one by one, whereas the paper-based questions were all printed on one page. The other was that in the computer-based version, each option was a selectable button as Figure 20 shows:

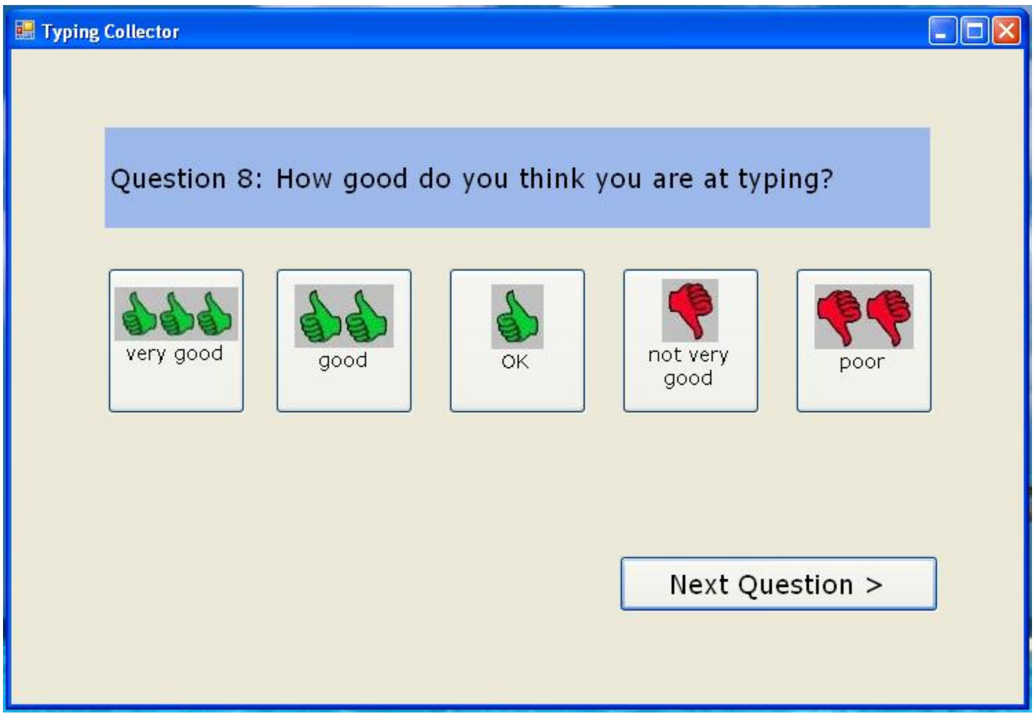


Figure 20: Thumbs-Up Scale Items in the TypingCollector shown as Buttons with Images

7.7.4 Procedure

The children attended a morning event organised at the author's department, taking part in various research activities. At the start of the day, the children were selected by their teacher to be in groups of four. Several activities were available in separate rooms and the children took turns to carry out all of the tasks. The questionnaire activity was carried out in a quiet computer lab with four children (1 group) at a time. Each group was assigned one of two conditions - paper followed by computer, or computer followed by paper.

As the children entered the room, they were asked to sit in front of a computer individually. There were four identical PCs available in a line. The PCs sat on individual desks that had ample room to also fill in the paper-based questionnaire.

If the group had the condition of computer-then-paper, the children were asked to follow the instructions given by the TypingCollector on their screen. Once a child completed the TypingCollector, they were then asked to fill in the paper-based questionnaire. The groups that had paper-then-computer condition were asked to carry out the reverse. Each child completed both modes of questionnaire at their individual desks.

7.7.5 Design

The digitised version of the six CCTEQs was administered through the TypingCollector. The TypingCollector asked some demographic questions (gender, age, handedness, etc), then asked the CCTEQs. The participants were then asked to copy type 10 phrases as before.

The paper-based questionnaires contained only the six CCTEQs across one page, written in Comic Sans font size 12pt. These questions were identical to the CCTEQs asked in the computer-based questionnaires, in order and appearance.

7.7.6 Analysis

To measure the correlation between the two modes, the answers the children gave to the rating scales were coded numerically. For the frequency of computer use at home, the answers were coded as 1 for '0 days', 2 for '1 to 2 days', 3 for '3 to 4 days', 4 for '5 to 6 days' and 5 for '7 days'. For the questions regarding how good and fast they think they are at typing, the answers were coded as 1 for 'poor/very slow', 2 for 'not very good/slow', 3 for 'okay/average', 4 for 'good/fast' and 5 for 'very good/very fast'. Answers to the question asking how much they liked typing were coded as 1 for 'I hate it', 2 for 'I don't like it', 3 for 'I don't mind', 4 for 'I like it' and 5 for 'I like it very much'.

The participants' typing speed and accuracy were calculated in the same manner as the previous studies.

7.7.7 Results

Each condition (paper-then-computer and computer-then-paper) had equal numbers of boys and girls, with six boys and four girls in each. Completion rates on both paper and computer based questionnaires were 100%.

7.7.7.1 Order Effect

To evaluate for order effects of paper-then-computer and computer-then paper, two Mann-Whitney U tests were performed to compare all of the questions. The Mann-Whitney test was chosen over the Independent Sample t-test since with the exception of AgeFirstUsed and PcInBedroom, all other variables were ordinal. Table 31 presents the results of the order effect for answers given in the paper-based questionnaire, and Table 32 presents the same range of information for the computer-based questionnaire.

Table 31: Scores on the Paper-based Questionnaire - Median, *U* Value, *Z* and *p* Value (Two-Tail, N = 20)

CCTEQ	Median - Paper first	Median - Paper second	*U*	*Z*	*p*
AgeFirstUsed	5.00	4.50	47.0	-0.238	0.812
PcInBedroom	1.00	1.00	45.0	-0.457	0.648
FreqUse	5.00	3.00	25.0	-1.977	0.048
TypingSkill	4.50	3.50	33.5	-1.305	0.192
TypingSpeed	3.00	3.00	41.5	-0.671	0.502
TypingAttitude	4.00	4.00	39.0	-0.857	0.391

Table 32: Scores on the Computer-based Questionnaire - Median, *U* Value, *Z* and *p* Value (Two-Tail, N = 20)

CCTEQ	Median - Computer first	Median - Computer second	*U*	*Z*	*p*
AgeFirstUsed	5.00	4.50	43.0	-0.560	0.575

PcInBedroom	1.00	1.00	50.0	0.000	1.00
FreqUse	5.00	3.00	27.0	.-1.826	0.068
TypingSkill	4.00	3.00	32.0	-1.442	0.149
TypingSpeed	3.00	3.00	45.0	-0.393	0.694
TypingAttitude	4.00	4.00	41.5	-0.665	0.506

With the exception of one question, no significant order effect was observed. The only question that did have a significant order effect was the frequency of use at home when answered on paper.

7.7.7.2 Format Effect

To evaluate the effect of the two forms of administration, the variables were merged together into answers given on the computer and paper, regardless of the order. The Wilcoxon Signed-Rank test was carried out. Table 33 shows the significance for each of the questions for each mode of administration of the questionnaires (paper and computer-based).

Table 33: Correlation between Paper and Computer-based Questionnaire Answers - Median, *Z* Value, *p* Value and *r* (N=20)

CTEQ	Median - Paper	Median - Computer	*Z*	*p*	*r*
AgeFirstUsed	5.00	5.00	-1.00	0.317	-0.16
PcInBedroom	1.00	1.00	-1.00	0.317	-0.16
FreqUse	3.50	3.00	-1.414	0.157	-0.22
TypingSkill	4.00	3.00	-.966	0.334	-0.15
TypingSpeed	3.00	3.00	-1.00	0.317	-0.16

For all the questions, no significant ($p>0.05$) format effect was found between paper and computer-based questionnaires.

7.7.8 Discussion

The study showed that the format of the questionnaire between paper and computer did not have any significant effect in young children. This suggests that children are able to answer questions regarding their computer and typing experience just as effectively on computer as they do on paper.

There was one case where the order of presentation of format had a significant difference in the answer the participants gave. There was a significant difference between paper-then-computer (Md = 5.00) and computer-then-paper (Md =

3.00) for frequency of use of computers at home on paper (U = 25.0, p = .048, r = 0.44).

In terms of individual cases, out of the 12 mismatched pairs between answers given on paper and computer versions, most only varied by one scale unit. Only two cases had differences of more than two scale units. This suggests that although there were mismatches, children gave similar answers across the two modes, rather than markedly different answers.

One participant (C168) produced four mismatches when most other participants only had one mismatch. It is suggested that a threshold of number of mismatches per participants should be set, and perhaps that this participant may be removed from the sample set. Another participant (C171) stated remarkably different ages for firstUse between the paper (9 years old) and computer (5 years old). This error probably came about due to the layout of the paper-based questionnaire. The paper version simply contained the six CTEQs. Therefore, the first question the participant saw was 'How old were you when you first used a computer keyboard?' Since it is often the case in forms that they are asked how old they are, it is possible that C171 misread the question as asking for his age. This assumption is supported by the fact that the participant's real age was indeed nine years.

The paper-based questionnaire suffered from some unclear answers. For instance, C177 selected their answer for each question by crossing out the one they wanted to select. The crossing out is easy to interpret as a selection for the rating scale, but not so easy for the yes/no answer. Other difficulties encountered were ambiguous numbers given in the firstUse age, and crossing out of answers when the participants changed their minds. In contrast, there were no ambiguous answers given in the computer version, since the participants were only able to select one option at a time, in a clear manner. This finding is consistent with similar studies with adult participants (Petitt, 2002).

7.7.9 Conclusions

This study showed that children as young as eight years old are able to answer a computer-based questionnaire just as accurately as paper-based questionnaire.

This implies that it is possible to use computer-based questionnaires in this current work to collect CTEQ from children. This will allow for an integrated TypingCollector that will collect the demographic data, CTEQ data and phrase-copying task data all together, reducing the workload of the researchers running phrase-copying studies with children.

Since the integrated TypingCollector can collect all necessary data automatically, the investigators are no longer required to keep a close eye on each participant and make sure that the paperwork does not get mixed up. It will be possible for investigators to run the study on many participants in one go. The integrated TypingCollector will also allow for the possibility of sending out copies to schools rather than the investigators attending each data collection.

7.7.9.1 Limitations

One of the limitations of this study was the sample size and method of sample selection. Only 20 participants were studied, which possibly caused the anomaly in the order effect for frequency of use of computers at home. Only two participants gave mismatched answers for the frequency of use, but it is possible that the small sample number magnified the difference. It is acknowledged that further work needs to be carried out on a larger sample set. This larger sample set should also be collected from different classes and schools, as a sample collected from one class in one school is not representative of the population.

The questionnaire used in this study did not cover all aspects of CE. Further work should be carried out to design questions to measure all four aspects of OCE (amount of computer use, opportunity to use computers, diversity of experience, sources of information) and three aspects of SCE (perceived competency, control and perceived usefulness). This work will lead onto quantifying a unified, weighted number for each child's CE.

7.7.9.2 Contributions

This study contributes the knowledge that children have no difficulty in answering questions about their CE on a computer-based questionnaire. In addition, in asking children about their computer experience, the mode of administration (paper or computer) can be used interchangeably.

7.8 STUDY 4 - VALIDATING THE CHILD CTEQ

This final study for the current chapter focuses on validating the child questionnaire. The six child computer and typing experience questions (CCTEQs) were used to gather CE data from children, which were then tested for variability.

7.8.1 Method

A three-day study collected CE data from 48 children from two local primary schools. The digital version of the child questionnaire described in the previous study (Study 3, Section 7.7) that had been built into the TypingCollector was used to administer the questions. Scores of CE were tested for their variability within the sample to see whether or not they were normally distributed.

7.8.2 Participants

In total, 48 children from two schools took part in this study. They were all aged between 8 and 11 (Years 4 to 5). 27 were boys and 21 were girls. They are identified as group 12 and 13 (participants IDs C184-C231) in the participant summary chart found in Appendix 3 and represents children from two different schools. The second school (26 children, 17 boys, 9 girls, aged between 8 and 11 years) were taking part in an after-school computing club run by their IT teacher.

7.8.3 Apparatus

The child questionnaire consisted of the same six CCTEQs that were used in Study 3 (Section 7.5). In summary, the questions asked were:

- CCTEQ1: *How old were you when you first used a computer keyboard?*
- CCTEQ2: *Do you have a computer or a laptop in your bedroom?*
- CCTEQ3: *How many days a week do you use a computer at home?*
- CCTEQ4: *How good do you think you are at typing?*
- CCTEQ5: *How fast can you type?*
- CCTEQ6: *How much do you like typing?*

The CCTEQs used the Frequency of Use Scale (FUS - Section 7.5.1.2) for CCTEQ3 and Thumbs Up Scale (TUS - Section 7.5.1.1) in questions CCTEQ4 to 5. The questions were administered through the TypingCollector as before.

7.8.4 Procedure

One school class of 22 school children (group 12) attended a morning event organised at the author's department, taking part in various research activities. At the start of the day, the children were selected by their teacher to be in groups of four. Several activities were available in separate rooms and the children took turns to carry out all of the tasks. The questionnaire activity was carried out in a computer lab with two other small activities. Groups of eight to ten children were brought into the room at a time. A pair of children carried out the task at a time, with all children taking part in the task by the end.

For the second school, the author attended after-school computer club sessions at the school across two days. The teacher of the computer club first introduced the author to the pupils. The author then carefully explained the purpose and the procedure of the task before asking the children to carry out the task on a voluntary basis.

In both instances, the children were asked to answer demographic questions, the six CCTEQs and then carry out a ten phrase-copying task. The entire procedure was administered through TypingCollector and the author only assisted in the task if there were any problems.

7.8.5 Analysis

Five of the questions contained a 5-point rating scale, so the answers were scored from 5 to 1, with 5 being the most positive responses (e.g. 'very good', 'very fast', 'very useful') and 1 being the least positive responses (e.g. 'very bad', 'very slow' 'very useless'). A simple CE score was calculated for participants by totalling their score for all six questions.

7.8.6 Results

Figure 21 below shows that the total CE score for the children sample was normally distributed (Kolmogorov-Smirnov test, D (48) = 0.033, p > 0.150).

Table 34 below also shows that answers to all five CCTEQ questions with 5-point rating scale were normally distributed.

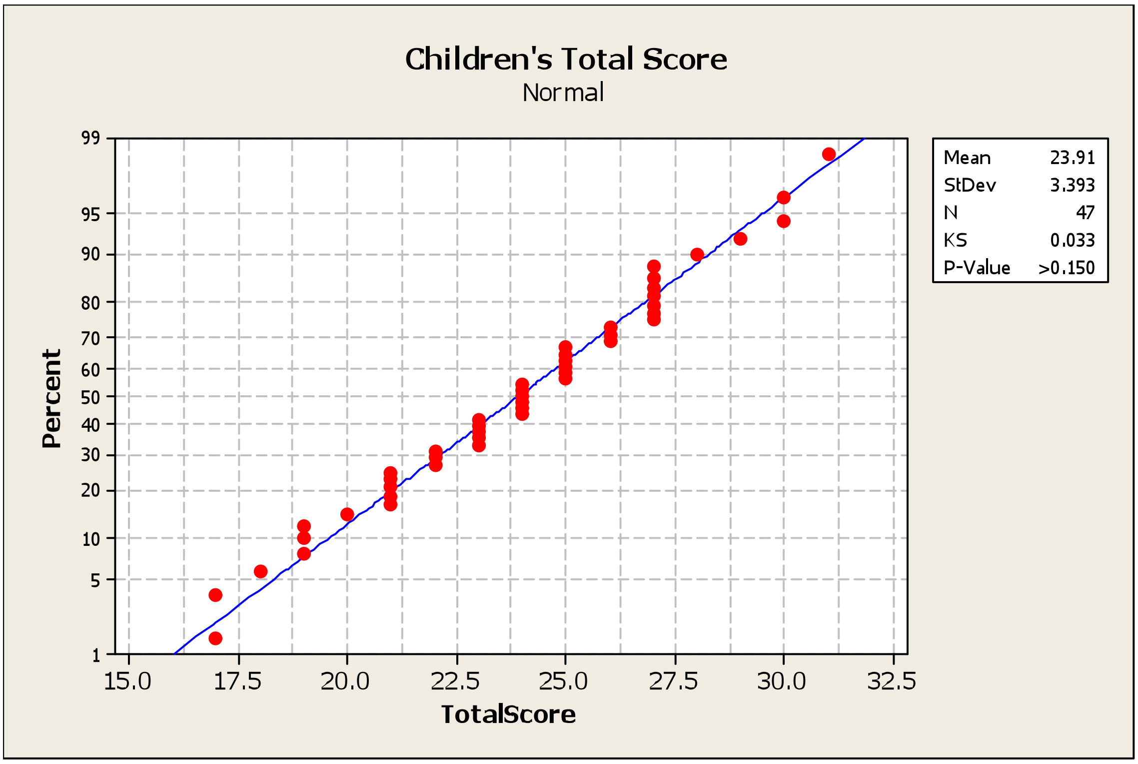


Figure 21: Normal Q-Q Plot for Total CE Score for Children

Table 34: Results of Kolmogorov-Smirnov Test on Individual CCTEQs

	Item	Mean	*D*(48)	*p*
CCTEQ 1	YearsOfComputerUse	4.426	0.055	> 0.150
CCTEQ 2	FreqUseAtHome	3.404	0.072	> 0.150
CCTEQ 3	CompetenceTypingSkill	3.574	0.042	> 0.150
CCTEQ 4	CompetenceTypingSpeed	3.000	0.036	> 0.150
CCTEQ 5	AttitudeToTyping	4.000	0.049	> 0.150

7.8.7 Discussion

For this particular sample of children selected, their overall CE score and the answers to each specific item on the questionnaire were normally distributed. This is in line with what was expected as there were no known demographical skew regarding CE such as those that were expected from the first-year computing students in study 1 (Section 7.4).

7.8.8 Conclusions

Using the Child Computer and Typing Experience Questions, it was possible to show that the child participant in this particular study showed normally distributed CE, with no demographical biases. However, this finding does not have the same validity as having a normal distribution in all 19 Adult CTEQs since the children's questionnaire only consisted of five questions. It is suggested that further work should be carried out to evaluate how many questions regarding CE the children will comfortably be able to answer (i.e. without getting bored), and maximise the questions to capture as wide a range of CE as possible within this.

7.8.8.1 Limitations

One important quality of any questionnaire is its reliability - if the same group of participants completes the questionnaire twice within a relatively small timeframe, then the answers the questionnaire extracts should be more or less the same. Neither the adult nor child questionnaires have been tested for reliability. A further study should be carried out to measure the reliability of the questionnaires.

7.8.8.2 Contributions

This study defined a new questionnaire for children to gather their previous CE. Examining the answers given to each question in the questionnaire can measure where some of the bias from the sampling method lies in the sample's CE.

7.9 CONCLUSIONS

This chapter has explored the methods in which to gather computer experience from both adults and children.

The chapter first surveyed the literature for various questions used to gather people's CE. It found that there were further subcategories to the established four OCE and three SCE categories. It was also found that questions regarding range of tasks and software used were outdated for use, and that very few questions were directed to typing itself.

In this chapter, two questionnaires were developed - the Adult Computer and Typing Experience Questionnaire (ACTEQ) and the Child Computer and Typing Experience Questionnaire (CCTEQ).

The first study in this chapter presented a group of first-year undergraduate computing students a question from each of the subcategories of CE. Some of the questions were updated with modern tasks and software, and others were changed to ask specific questions about typing. The distribution of the answers given to each questions were tested for normality and some were found to be positively skewed. The ACTEQ provides a more detailed understanding of where the demographic of the sample affects their CE.

The second study investigated whether or not children were able to accurately answer questions regarding their CE. It found that children were able to answer questions regarding their frequency of use, amount of use, perceived competence and attitude well. However, they experienced problems answering questions regarding particular software. It was recommended from this study that asking young children about range of tasks and software would yield inaccurate answers if they did not know the specific names of each piece of software.

The third study investigated if it was possible for children to answer CE questions on a computer, rather than on paper, and still give similar answers. It was found that most children gave the same answer, whether it was on paper or on computer. This suggests that it is acceptable to ask CE questions to children using the TypingCollector, without affecting the answers they give.

The fourth study defined and evaluated six questions regarding children's CE. It was found that for the sample used in the study, no aspect of the CE measured were skewed in any way and their total CE scores were normally distributed.

All four data collecting issues highlighted in the literature review (Chapter 2) have now been addressed. The data collection method for this research is a screen-to-screen phrase-copying task, using short phrases that have been designed for children. Adult participants and child participants have separate questionnaires. The TypingCollector administers all data collection tasks.

The next two chapters address the issues raised regarding the analysis method that is comparable for both children and adults.

8 CATEGORISATION OF TYPING ERRORS

8.1 INTRODUCTION

Chapter 3 highlighted the need for further investigation of the suitability of categorisation methods for use with children. This chapter focuses on the reduction of bias in text input studies between adults and children by addressing the analysis method.

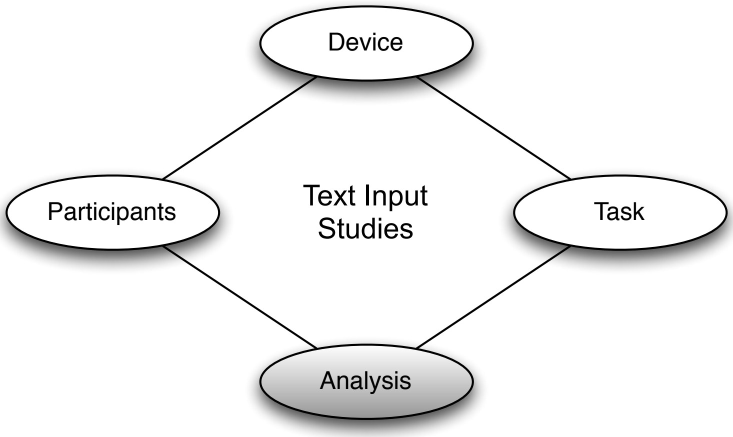


Figure 22: This Chapter Focuses on Reducing Bias in the Analysis Method

Initially, two existing categorisation methods were evaluated for their effectiveness in capturing all the typing errors made by children. The methods were applied to typing errors collected from an empirical study with children carrying out a text copy exercise. The methods were compared with visual inspection of the typing, and were found to either ignore many of the typing errors, or inflated the total number of typing errors by breaking larger errors into several smaller errors (error inflation).

In light of this, a new categorisation method was defined that captured all the typing errors displayed by the children in this study. Selections of error types defined in previous works have also been included. However, care was taken to ensure that the error types did not imply a cause, such as 'spelling error' (Logan, 1999) or assume formal touch-typing training, such as 'using the wrong finger' (Opfer, 1932). The new categorisation method was able to classify almost all of the errors collected from the study (2308 out of 2312 errors). Of the 2308 that were classified, only 19 were ambiguous. However, manually using this categorisation method was laborious and suffered from some inter-rater errors.

The work reported in this chapter was published at British HCI 2007 (Kano et al., 2007).

8.1.1 Objectives

The objective of this chapter was to:

1. *Evaluate existing typing error categorisation methods with real typing data collected from children*

   Section 8.2 describes a study carried out with children to collect real typing data. Existing categorisation methods were applied to the real data to see how efficiently they captured the typing errors.

2. *Design a more efficient categorisation method that captures all typing errors without suffering from error inflations or assuming causes*

   Section 8.4 introduces a new categorisation method that combines the typing errors found in children's typing, with those defined in previous methods to capture typing errors more fully and reduce error inflations.

8.1.2 Scope

Since the scope of the thesis is set within the full-size QWERTY keyboard, this was the only keyboard used in the data collection. It is acknowledged that other error types exist in other text input devices.

Similarly, since the scope of the thesis is set within the English language, only English speaking children took part. It is recognised that different typing errors may exist in different languages.

A decision was made to reuse the data collected in the two pilot studies previously described in Chapters 2 and 3. Although children experienced some difficulties with the paper-to-screen, it was felt that the data was still valid. It also captured many typing errors (such as omitted phrases) that children do make, but are unlikely if the phrases were shown to them one at a time.

8.1.3 Contribution

The main contribution of this chapter is the new categorisation method defined in Section 8.4. The categorisation method is thorough and robust against children's typing. It is also the first method that is based on real typing data collected from children.

8.1.4 Structure

Section 8.2 describes the empirical study carried out to gather the typing data from children, and how the existing categorisation methods were applied to the collected data. Section 8.3 outlines the evaluation results of the existing methods. A new categorisation method is introduced in Section 8.4, with definitions of error types for letter-level errors defined in Section 8.5, word-level errors in Section 8.6, phrase-level errors in Section 8.7 and others in Section 8.8. Section 8.9 applies this new categorisation method to the empirical data. The chapter concludes in Section 8.10.

8.2 METHOD

Data collected from the first two pilot tests (described in Chapters 2 and 3) were used for this study. The two pilot tests combine to make a dataset of 112 children (Children group 1 and 2 in Appendix 3) from two local primary schools to create a large amount of phrase copying data for analysis.

8.2.1 Participants and Apparatus

57 boys and 55 girls, aged between 5 and 10 years from two local primary schools took part. The study was carried out in a quiet room in both schools, using four identical black keyboards (PC Line PCL-K350) connected to four identical tablet PCs (RM Tablet PC CE0984) on stands (not used as tablets, simply used to create a consistent display) as Figure 23 shows:

Figure 23: Set Up of the Experiment

8.2.2 Design

Children copied phrases shown on paper into Notepad™ via a standard QWERTY keyboard. This was to enable the capture of all possible typing errors. Based on findings of previous studies (Kano et al., 2006), younger children were only asked to type five phrases each, while the older children typed ten. They were all given half an hour to complete the task.

The phrases to be copied were chosen by randomly selecting 50 phrases each from two phrase sets, TEPS (MacKenzie, 2006) and CPSet (Chapter 5). The 100 resulting phrases were each presented approximately 10.6 times. The phrases shown were randomized to eliminate learning effects. In all, 1060 phrases were shown to the children.

8.2.3 Procedure

Participants were selected by their teachers using guidance from the researchers to ensure a representative sample, across age and gender. The children came, in fours, to the room voluntarily and sat in front of an individual tablet PC/keyboard. The procedure was explained to them individually and three researchers oversaw the study.

Each child was given its own set of phrases on a sheet of paper in 20 point Arial. The children were instructed to type the phrases using the keyboard and were advised that the trial was not timed, nor marked. During the trial, every keystroke was recorded using KGB Keylogger®; this gave us an Input Stream (IS) that included all typed characters and other key presses, whether or not they appeared in the final text. Once the child completed the task, he or she left the room and was replaced by another child.

8.2.4 Analysis Method

8.2.4.1 Manual Classification

Firstly a manual analysis of the data was carried out to gauge the total number of errors made by the participants. Although manual analysis of the input stream is not 100% reliable, it allows the flexibility of highlighting all errors without the bias of categorisation and provides us with a point of comparison for other methods.

The PT was compared with the complete IS. Any errors found were noted for whether or not an attempt at fixing the error was made, as shown in Figure 24 below:

```
PT: back to my home
IS: bach<k to m house<<<me
```

Figure 24: An example PT (Presented Text) and IS (Input Stream)

In this example, where a Backspace is indicated by '<', there are a variety of errors. The word '`back`' was spelt with an '`h`' which was then Backspaced to fix it, so this would count as one corrected error. Another is an uncorrected error where '`my`' was typed without the '`y`'. The word '`home`' was typed as '`house`' and this is considered, in this classification, as one corrected word error.

8.2.4.2 Gentner et al. Classification

A second visual classification was then carried out based on the eight error types defined by Gentner et al. (1983). This used only TT ignoring the IS. The classifications were Transposition, Interchange, Migration, Omission, Insertion, Substitution, Doubling Error and Alternating Error. Figure 25 below shows TT from the previous example:

```
PT: back to my home
IS: bach<k to m house<<<me
TT: back to m home
```

Figure 25: Calculated TT (Transcribed Text) From the Given IS

There is only one error remaining in the TT: the omission of '`y`' in '`my`'. This method does not count any errors that were corrected in Input Stream and thus has lower number of errors than the manual classification method.

8.2.4.3 Wobbrock and Myers Classification

The third method carried out, classified errors as defined in Wobbrock and Myers (2006): Corrected Omission, Uncorrected Omission, Corrected Insertion, Uncorrected Insertion, Corrected Substitution, Uncorrected Substitutions, and Corrected No-Error.

As the children had not used the Wobbrock and Myers (2006) StreamAnalyser, the researcher had to input the IS along with the PT into the StreamAnalyser. When there were several possible combinations of error types, the program offered all possible combinations and the researcher chose the least costly combination.

8.3 RESULTS

8.3.1 Manual Inspection of Key Logs

Table 35 below shows that out of the 1060 phrases shown, the children omitted 30 phrases, attempted to copy 1030 phrases (25531 letters), of which they left 7 incomplete. In total, 2312 errors were found. 125 were word-level errors and the remaining 2187 were letter-level errors.

Table 35: Result Summary of Manual Inspection

Total No. of Errors found	**2312 errors**
Letter errors	2186 errors
Word errors	125 errors
No. of fixable errors	**2290 errors**
No. of corrected fixable errors	1132 errors
Corrected immediately	714 errors
Corrected by backspacing	300 errors
Corrected by left/right keys	118 errors
No. of uncorrected fixable errors	1158 errors

Of the 2312 errors, 99% were fixable errors (e.g. pressing a function error was not fixable). However, only half of the errors (49.4%) were fixed. Of these corrected errors, 63.1% of the time the child noticed the error immediately and thus only required a single Backspace to delete the erroneous letter. Of the remaining 418 errors that were only noticed after typing a few more letters, 300 errors were reached by backspacing all the letters in between and 118 errors were reached by pressing the Left and Right directional keys.

8.3.2 Gentner et al. Classification

This categorisation method identified 1327 errors in the transcribed text. Table 36 below shows the most common errors were Insertion (571 errors, 43%), followed by Omission (443 errors, 33.4%) and Substitution (300 errors, 22.6%). The remaining errors accounted for less than 1% of the overall categorised errors.

Table 36: Summary of Gentner et al. (1983) Classification

	Frequency	% of total error
Doubling Error	2	0.15
Insertion	571	43.03
Migration	2	0.15
Omission	443	33.38
Substitution	300	22.61
Transposition	9	0.68
Alternating	0	0.00
Total	1327	

8.3.3 Wobbrock and Myers Classification

This analysis found 2490 errors in total. A summary of the counts is shown in Table 37.

Table 37: Error Counts Found by Wobbrock and Myers (2006) Classification

	Corrected	Uncorrected	Total (% of total number of errors = 2490)
Insertion	403	555	958 (38.5%)
Omission	238	518	756 (30.4%)
Substitution	433	343	776(31.13%)
Total (% of total number of errors = 2490)	1074 (43.13%)	1416 (56.87%)	2490

There was a more even spread of errors among Substitution, Omission and Insertion than in the Gentner classification. Compared to the manual classification, the Wobbrock and Myers analyser reported more uncorrected errors, principally because, due to the design of the analyser, corrections involving the Left and Right directional keys along with those made after the Return key was pressed at the end of the phrase, were omitted.

This analyser also classified 391 occurrences of Corrected No-Error but did not distinguish between those that occurred during editing to those erroneously erased. Corrected No-Errors were counted by number of occurrences, i.e. if several correct letters were deleted in a row, they were counted as one Corrected No-Error.

8.3.4 Comparison of the Categorisation Methods

The categorisation methods defined by Gentner et al. (1983) and Wobbrock and Myers (2006) were compared with the manual inspection, to find out to what extent each method ruled out certain error types, and how the different definitions of errors affected the result. Table 38 summarises the number of errors found by each method.

Table 38: Comparison of the Three Categorisation Methods

	Manual	Gentner	Wobbrock
Total No. of errors classified	2312	1327	2490
No. of manually found errors left unclassified	-	1199	203
% of manually found errors left unclassified (2313)	-	51.9%	8.8%

8.3.4.1 Gentner et al. Classification

Of the 2312 errors found in the manual inspection, 1199 (51.9%) did not feature in Gentner's classification as they were fixed during the experiment by the participants and thus did not remain in TT. In addition, due to the character-based nature of the definitions of error types in (Gentner et al., 1983), some errors found manually, such as words omitted or inserted, were regarded as multiple single-letter errors. This occurred 81 times, with varying number of errors produced in each.

8.3.4.2 Wobbrock and Myers Classification

Out of the 2312 errors found manually, only 203 errors (8.8%) were left unclassified by Wobbrock and Myers' classification. Most of the unclassified errors were made on the phrases after the Return key was hit at the end of the phrase. These were not counted as errors due to the fact that, if the children were carrying out the copy task with Wobbrock and Myers' accompanying program TextTest, they would not be able enter any text for the phrase once the Return key was pressed. For the same reason, errors and fixes made during Left and Right arrow moves were discounted as the keys are disabled in the TextTest program. Errors caused when a function key was pressed unintentionally, or hitting the Return key in the middle of a phrase and carrying on typing on a new line, which was allowed in our study, were also not included in the total count of errors. As with Gentner's classification, due to the character-level nature of the error type definitions, an individual word-level error in the manual classification was considered as multiple single-letter errors.

This classification method does not differentiate between a correctly typed character deleted during editing and one deleted by error. If this distinction between the two types of error is made, 114 out of the 391 Corrected No-Error occasions are identified as erroneous. Wobbrock and Myers (2006) categorisation stripped the error types down to the bare minimum of Insertion, Omission and Substitution. This means that the more complex errors, such as Migration, Alternating Error and Interchanges are also broken down into multiple, simpler error types. For example, a Migration error, such as shown below where the first c migrates across three letters:

`orrcection (correction)`

is classified as Omission of the first `c`, then an Insertion of the `c` later in the word.

Both methods fail to capture the entire range of errors for a phrase-copying task using a QWERTY keyboard. Gentner et al. (1983) categorises typing errors into more detailed error types, but applying this to only the transcribed text reduces the number of classified errors dramatically. In contrast, Wobbrock and Myers' (2006) categorise more errors by considering those corrected in IS. However, it loses the detail of what errors were made due to featuring a reduced number of error types.

8.4 ExpECT - A New Error Categorisation Method

Although both Gentner et al. (1983) and Wobbrock and Myers' (2006) categorisations work well when comparing different text input methods on one user group, they are not ideal for capturing all the errors made during a phrase-copying task, especially when the participants are children.

To carry out a thorough comparison between the way children and adults made typing errors, it was necessary to define a new categorisation method. The requirements of the new categorisation method were:

1. Be able to capture as many typing errors as possible by a set of well-defined error types

2. Limit error inflation
3. Not assume participants have had any formal typing training
4. Error types must not make assumptions as to the cause of the typing errors

In this new method, error types in the literature were merged with several new error types not previously defined. Most significantly, White's (1932) categorisation method has been extended by introducing more word level errors and even phrase level errors. These error types, which are concerned with errors on a larger scale, will provide a 'bigger picture' of what goes wrong in a copy-typing task.

'Next To' and 'Close Errors' (Read and Horton, 2006) have been separated into those that cause a substitution of the intended letter (NT-S and CT-S) and those that cause a multiple number of characters to be inserted (NT-Mu and CT-Mu). Close Errors, defined in (Read and Horton, 2006) has not been altered in its definition, but the name has been changed to Close To (CT) for simplicity, also reflecting its similarity in properties with NT errors.

Substituted and inserted words were defined into two types according to the source of the new word. Children often lose their place on a sheet of writing while reading or copying onto paper and find their place again elsewhere on the sheet, or replace words with those that are similar in context.

Finally, Corrected No-Errors that occur erroneously, and those erased intentionally during editing are differentiated. Only those deleted erroneously are considered as an error in this method.

Although some of the error types defined here may seem unlikely for experienced adult typists, the earlier study in this chapter suggests that they occur frequently during text input with young children.

The error types defined here are grouped according to the levels of detail they are concerned with, either at a letter, word, or phrase level as shown in Table 39:

Table 39: Summary of the New Error Types

	Omission	Substitution	Insertion	Other
Letter	OL OS	AE CaE CT-S DE IE NT-S SL TE ME	CT-Mu DL DS IF IL IS ISy NT-Mu	U ExE CNE(error)
Word	OW	SW-A SW-U	IW-A IW-U DW	
Phrase	OP	SP	DP EE	

8.5 LETTER-LEVEL ERRORS

Letter-level errors are the most common mistakes in typing for both adults and children and thus have the most extensive range of error types to categorise into.

8.5.1 Omission Error Types

8.5.1.1 Omitted Letter (OL)

When a letter is omitted from the word when it is typed, it is classified as an OL (Omitted Letter) error. Some examples are (intended text in brackets):

```
litte (little)
brething (breathing)
```

8.5.1.2 Omitted Space (OS)

In an OS (Omitted Space) error, a space is omitted from a word where there should be one according to the intended text. Some examples of OS errors are shown below:

```
thanksfor (thanks for)
doorsare (doors are)
```

8.5.2 Substitution Error Types

8.5.2.1 Substituted Letter (SL)

An error is classified as an SL (Substituted Letter) error when an incorrect letter substitutes the intended letter and it cannot be classified as any other letter-level substitution error types (TE, NT-S, CT-S, AE, IE, ME). Some examples of SL errors are:

```
fiowers (flowers)
rounp (round)
```

8.5.2.2 Transposition Error (TE)

The definition of Transposition errors (TE) remains unaltered from Gentner et al. (1983) '*when consecutive letters are switched. Also occurs when space or punctuation that precedes or follows the word is switched.*' Some examples of this error type are:

```
littel (little)
tiem (time)
```

8.5.2.3 Next To error – Substitution (NT-S)

An error is classified as an NT-S (Next To error – Substitution) when a key directly next to the intended key is pressed, producing a different letter instead of the intended letter. Some examples of NT-S are:

```
thinga (things)
a;ways (always)
```

NT and CT keys are dependent on the keyboard layout (in this case QWERTY) and also on the particular model of the keyboard. Figure 26 shows that, if the intended key is 'G', then the keys 'F' and 'H' are classified as NT keys:

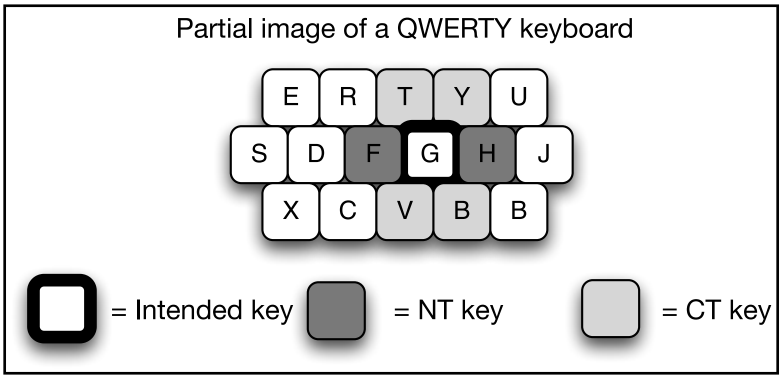


Figure 26: NT and CT Keys for 'G' on a QWERTY Keyboard

Although in some cases of NT-S it is possible that it was a spelling mistake (SL error), any error where the intended letter is substituted with an NT letter is classified as an NT-S error.

8.5.2.4 Close To error – Substitution (CT-S)

'Close To' keys to an intended key are those keys neighbouring the intended key, either above or below it. It is possible to press a Close To key accidentally instead of, or together with, the intended key. Figure 26 shows a partial layout of a keyboard highlighting the CT keys for the key 'G'.

In a CT-S (Close To error – Substitution), a intended letter is substituted by a CT letter. Some examples of CT-S errors are:

```
goldeh (golden)
tye (the)
```

As with NT-S, although it is possible that a CT-S error could actually be a SL error where there was a spelling mistake, any letter substitution where the intended letter was substituted by a CT letter is classified as a CT-S error.

8.5.2.5 Capitalisation Error (CaE)

When either a capital letter in the presented text is typed as a lower case letter, or vice versa, it is classified as a Capitalisation Error.

8.5.2.6 Alternating Error (AE)

The definition for Alternating Error (AE) remains unchanged from Gentner et al. (1983): '*when a letter alternates with another but the wrong alternation sequence is produced*'.

```
threr (there)
```

AE errors are restricted to those words where the intended word contains a three-letter combination of the first and last letter being the same character.

8.5.2.7 Doubling Error (DE)

The definition for a Doubling Error (DE) remains unaltered from Gentner et al. (1983), '*word containing a repeated letter and the wrong letter is doubled instead*'.

```
caleed (called)
```

Doubling errors are restricted to those words where the intended word contained two consecutive letters, which are the same. If a single letter is duplicated, it is classified, as a Duplicated Letter instead.

8.5.2.8 Interchange Error (IE)

The definition of an Interchange Error (IE) across I letters remains unaltered from Gentner et al. (1983): *'two non-consecutive letters are switched with I letters intervening (I>0)'*.

8.5.2.9 Migration Error (ME)

The definition of a Migration Error (ME) across M letters remains unaltered from Gentner et al. (1983): '*one letter moves to a new position, with M letters intervening between its correct position and its end position (M>1)'*.

```
orrecection (correction)
```

8.5.3 Insertion Error Types

8.5.3.1 Inserted Letter (IL)

When an extra letter (not a duplicate of the previous letter) is inserted, it is classified as an IL (Inserted Letter) error. Some examples include:

```
hern (her)
docktor (doctor)
```

8.5.3.2 Duplicated Letter (DL)

When a character is erroneously repeated twice in a row, it is classified as a DL (Duplicated Letter) error. Some examples are:

```
alwaays (always)
appartments (apartments)
```

However, if the duplicated letter either precedes or follows an intentional double letter but was only typed once, it would be classified as a Doubling Error.

8.5.3.3 Next To error – Multiple key presses (NT-Mu)

When a key directly next to the intended key is pressed along with the intended key, producing the intended letter and one or more extra letters, it is classified as an NT-Mu (Next To error – Multiple key presses) error. Some examples of NT-Mu error are:

```
ourt (our)
agwes (ages)
```

As with the NT-S error, it is possible for an NT-Mu error to actually be another error type created from a spelling mistake (such as IL error). When the extra letter is an NT letter to the intended letter, it is prioritised to be classified as an NT-Mu error.

8.5.3.4 Close To error – Multiple key presses (CT-Mu)

When one or more keys 'close to' (see CT-S for definition) but not next to the intended key is pressed together with the intended key, producing the intended letter and one or more extra letters, it is classified as a CT-Mu (Close To error – Multiple key presses) error. Some examples are:

```
onl7y6 (only)
wr8i9ting (writing)
```

As with the CT-S error, it is possible for CT-Mu error to be another error type created from a spelling mistake (such as IL error), but when the extra letter is a CT letter to the intended letter, it is classified as a CT-Mu error.

8.5.3.5 Inserted Space (IS)

When an extra space is inserted where there should be no spaces according to the intended text, it is classified as an IS (Inserted Space) error. Some examples of this error type are:

```
t eam (team)
house keeper (housekeeper)
```

8.5.3.6 Duplicated Space (DS)

An error is classified as a DS (Duplicated Space) error if two spaces are typed when only one space is shown in the presented text.

```
all  that  he  could  see (all that he could see)
these  cookies (these cookies)
```

8.5.3.7 Inserted Symbol (ISy)

If a symbol is inserted when there are no symbols in the presented text, it is classified as an Inserted Symbol (ISy) error. As the phrase set used in our study contained no symbols, any symbols found were classified as an ISy.

8.5.3.8 Inserted Function (IF)

If a function key, such as Control or Alt is pressed when not required, it is classified as an Inserted Function error. IF errors are only found in the input stream since functional keys do not produce letters or symbols.

8.5.3.9 Corrected No-Error by Error (CNE(error))

Wobbrock and Myers (2006) refer to Corrected No-Errors as those letters that were correct but then are erased. There are two purposes for this action; one is to fix an error that is only a few letters from where the cursor is, deleting the letters in between. These are editing actions and therefore are not classified as errors. The other is deleting a letter either because the participant thought they made a mistake when they had not, or accidentally deleted a letter in the process of editing, by pressing the Backspace key too many times. These latter types are errors and are categorised as Corrected No-Error by Error or CNE(error).

8.6 Word Level Errors

Word level errors are less common than letter errors but do occur regularly with children. They are more likely to not read the presented text properly, or do not remember the exact words in the presented text, and alter the words in the phrase.

8.6.1 Omission Error Types

8.6.1.1 Omitted Word (OW)

In an OW (Omitted Word) error, an entire word is omitted while typing the intended phrase. Examples of this error are:

```
they all go marching (they all go marching down)
two one zero blast off (three two one zero blast off)
```

8.6.2 Substitution Error Types

8.6.2.1 Substituted Word – word from Another place (SW-A)

An error is classified as an SW-A (Substituted Word – word taken from Another place on the phrase sheet), when a word from the intended text is substituted by another word, and the substituting word is not one found within the phrase, but is found elsewhere on the phrase sheet. Examples of this error type are:

```
the etiquette (the objective)
she plays (she rules)
```

8.6.2.2 Substituted Word – word source Unknown (SW-U)

When a word from the intended text is substituted with another word and the substituted word does not appear in the phrase or the phrase sheet, the error is classified as SW-U (Substituted a Word – word source Unknown). Some examples of this error type are:

```
on my face (on his face)
back to my house (back to my home)
```

8.6.3 Insertion Error Types

8.6.3.1 Inserted Word – word from Another place (IW-A)

If an extra word is inserted, and the inserted word is not found within the same phrase but is found in another phrase on the phrase sheet, it is classified as an IW-A (Inserted Word – word found from Another place on the phrase sheet) error. An example of IW-A error is:

```
has been increased (has increased)
```

8.6.3.2 Inserted Word – word from Unknown source (IW-U)

When an extra word is inserted, and the inserted word is not found within the phrase or the phrase sheet, it is classified as an IW-U (Inserted Word – word from Unknown source) error.

8.6.3.3 Duplicated Word (DW)

A word can be duplicated within a phrase and is classified as a Duplicated Word (DW) error. DWs are restricted to errors where the duplicated words appear in the same phrase.

```
from the west the west (from the west)
```

8.7 PHRASE LEVEL ERRORS

Phrase level errors are more common in younger children who have difficulty keeping their place on a phrase sheet, and in remembering what they have just typed. Phrase level errors can be excluded if the participants are shown only one phrase at a time.

8.7.1 Omitted Phrase (OP)

When an entire phrase is omitted, it is classified as an Omitted Phrase (OP).

8.7.2 Substituted Phrase (SP)

An intended phrase may be replaced by another phrase on the phrase sheet. This error is classified as an SP (Substituted Phrase) error.

8.7.3 Duplicated Phrase (DP)

Duplicated Phrase errors (DP) can occur if a phrase that is already typed is typed again. This error is more prominent in younger children who cannot remember what they have already typed as well as older children can.

8.8 OTHER ERROR TYPES

8.8.1 Enter Error (EE)

When the Enter (Return) key is pressed in a place other than at the end of a phrase, it is classified as an EE (Enter Error). An example of EE is:

```
with bright shining
faces
(with bright shining faces)
```

8.8.2 Execution Error (ExE)

Defined by Read et al. (2001), Execution Errors (ExE) refer to those errors created by the person holding down a key for too long, resulting in multiple entries of the same letter, symbol or space.

```
maaaaany (many)
```

8.8.3 Unknown (U)

When the error does not fit into any of the above categories it is classified as a U (Unknown) error. An error is likely to be classified in this error type when it is difficult to guess why the error happened. This is different from an error with ambiguity of classification types, as that sort of error is easily understood as to how it was possible to create. Unknown errors are those where there is more than one possible way that the error was created, but typically no straightforward 'sensible' construction. Below is an example of a U error:

```
sgfsjkfjsimdj (???)
```

8.9 APPLYING THE NEW CLASSIFICATION

To test the validity of the newly defined error types, they were applied to the errors found manually from the earlier study of copying task carried out by children in Section 8.3.1. Table 40 shows a summary:

Table 40: Summary of Errors Categorised Using ExpECT

Error Type	Frequency	Error Type	Frequency
Ambiguous	19	IS	89
AE	0	ISy	170
CaE	290	IW-A	6
CNE(error)	129	IW-U	3
CT-Mu	30	ME	3
CT-S	23	NT-Mu	37
DE	4	NT-S	136
DL	64	OL	314
DP	4	OP	3
DS	195	OS	293
DW	15	OW	54
EE	28	SW-A	15
ExE	34	SW-U	16
IE	0	TE	24
IF	24	U (Unknown)	4
IL	107	Total	2312

With the new classification there were just 19 ambiguities. As IS was used as the source of the errors (as opposed to TT), many errors were noticed and fixed by the participants halfway through making the error. This made some errors difficult to guess without further knowledge about the intention of the participants. In particular, the ambiguity between OL with either NT-S or CT-S - when the substituting letter was the same as the letter following the intended letter - was difficult. In these cases, either NT-S or CT-S was chosen over other possible error classifications. Although this appears to be an assumption that would considerably alter the final result, such assumptions were made for only five of the errors found amongst the 2312 errors categorised. Nevertheless, to accurately solve these ambiguities further investigation into the intention of the participant when the error was created is required.

Identifying, classifying and recording the errors by hand was difficult. It is also a very costly method, as the errors had to be checked repeatedly to ensure a correct categorisation. However despite thorough checks by the first researcher, a second researcher categorised a portion of the errors found and noted that they disagreed upon 1.2% of the categorisation. An automated algorithm, which carries out these classifications, will be required for large studies to reduce the cost and time of the study and also to ensure that the errors are categorised correctly and accurately.

8.10 CONCLUSIONS

Since the early days of Lessenberry and Dvorak, understanding of how to measure text input methods in terms of errors has come a long way. Errors can be automatically categorised into well-defined sets of error types and speculate their causes.

This study applied two contrasting methods of categorising typing errors created by children during a copy-typing task. It was discovered that each method and sets of definitions lacked in some aspects of classifying all the errors found in such tasks. Some did not allow for several phrases to be shown at once, some only took errors that remained unfixed into the transcribed text, some only took few error types into consideration and ignored other error types, and others did not allow certain methods of fixing an error. A new typing error classification method was defined that combines previously defined error types with some new error types to create a more thorough and broader method in analysing typing errors.

Classifying errors by hand is difficult and not 100% accurate. An automated algorithm that carries out the categorisation will reduce cost and raise reliability in thoroughly investigating typing errors.

9 AUTOMATING CATEGORISATION OF TYPING ERRORS

9.1 INTRODUCTION

The study presented in the previous chapter (Chapter 8) required manual analysis of 106 participants typing 1030 phrases. This manual process took three weeks to complete. For a large-scale study, the process of analysing typing errors must be automated. Additionally, automation was desirable since it would ensure a systematic categorisation of typing errors, with disambiguation rules applied consistently. This would reduce the effect of the analysis method on the outcome of the study.

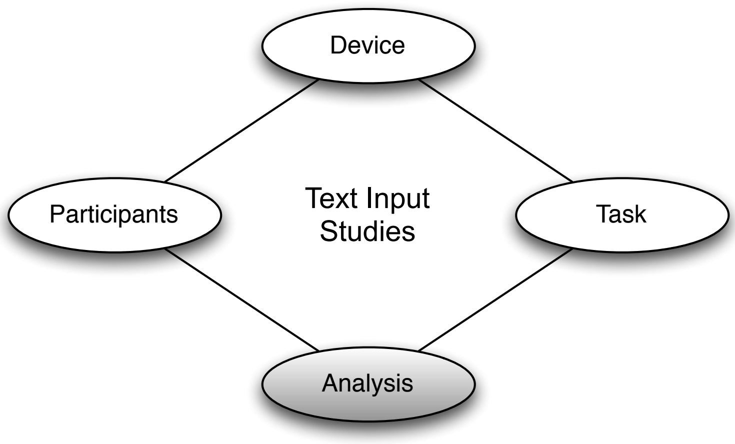


Figure 27: This Chapter Focuses on Reducing Bias form the Analysis Method by Ensuring Consistent Classification of Errors

This Chapter presents work carried out to *automate the categorisation of typing errors* in accordance with the categorisation method ExpECT. The new automated TypingAnalyser was evaluated with typing logs of 412 new participants. The result was manually checked and was found to be 99.6% accurate, and resolved 88% of ambiguous cases.

9.1.1 Objectives

The main objective for this work was *to create a program that would accurately categorise as many typing errors as possible*. To achieve the objective, this study had the following additional objectives:

- *To automate categorisation of typing errors*

  A program was required to automatically detect and categorise typing errors according to error types defined by ExpECT.

- *To define disambiguation rules in order to reduce ambiguities consistently*

  Automated categorisation of typing errors often resulted in more than one possible answer. Rules had to be defined on how to reduce these ambiguities in a consistent manner.

- *To reduce the time required to spend in formatting the data into something usable*

  Even after categorising all the typing errors found, researchers must spend a vast amount of time logging and extracting results before these can be analysed. Several additional features have been implemented in the TypingAnalyser and DataSummariser to reduce the effort required to prepare a summary of the data.

9.1.2 Scope

This study has several constraints. The TypingAnalyser is constrained to categorising typing errors that occurred during a phrase copy-typing task (see Section 2.3 for definition). It cannot analyse typing errors from a created typing task where no Presented Text (PT) is available.

The analyser is currently set to assume that the participant used a full size UK QWERTY keyboard (see Appendix 1 for a diagram of the layout). If another layout is used to collect the typing data, this must be reflected in the analyser by changing letter.java in the program.

The TypingAnalyser assumes that the presented text consists only of British English, and that the participants spell words with British spelling. If the analyser is to be used on American English speaking participants, specific words spelt differently between British and American English must be added to oxford3000.txt file.

At minimum, the analyser requires an xml format log of PT and IS as specified by the TypingCollector. A timestamp of each key press is also required.

The analyser assumes that all typing was carried out on a typing data collector (such as the TypingCollector used in this thesis) that only allowed pressing the Backspace key to edit what was typed. It does not account for editing done using the mouse or the directional arrow keys to move the cursor.

The study carried out to evaluate the accuracy of the analyser was constrained to 412 participants, totalling 3940 phrases. Although this contained both children and adults, it is acknowledged that it is not a large number of participants. The implications of this are discussed further in Section 9.10.3.2.

9.1.3 Contribution

The main contributions in this chapter are:

1. A Java program that can detect and classify typing errors accurately, with minimum error inflation or deflation.
2. Methods to classify word-level error types and further character-level error types that add contextual data to the typing errors found.
3. A set of rules to reduce ambiguities in classifying typing errors.
4. A second program that tabulates the results automatically in three different contexts - all the errors, all the phrases and all the participants.

9.1.4 Structure

Section 9.2 evaluates previous analysers, and identifies requirements for the new analyser. It also discusses the steps required for transforming raw typing data to tabulated results, and how the current work uses previous works as a foundation. Section 9.3 discusses how the new TypingAnalyser treats the Input Stream in a different way to previous works. Section 9.4 outlines methods used to detect character-level errors. Section 9.5 briefly describes how the error rates are calculated. This is followed by a discussion in Section 9.6 of how split character level errors are classified. Section 9.7 discusses how word-level errors are detected. Section 9.8 focuses on reducing duplicate and redundant results. Section 9.9 describes the steps taken to reduce the workload of the user in each step of the analysis. Section 9.10 and 9.11 describes the two studies carried out to evaluate the TypingAnalyser. Finally, Section 9.12 discusses the conclusions of this chapter.

9.2 MOTIVATION AND REQUIREMENTS

Manual categorisation of even small samples of typing is hugely time consuming. One must first compare PT and IS that are not visually aligned with each other. Figure 28 shows the raw output data that the researcher is faced with. It is easy to see why a manual search through raw data for errors is not only time consuming but prone to errors being misclassified, or not detected at all.


```
<PT>down by the pond</PT>
<IS>
<LETTER><TIME>633619249301801375</TIME><KEY><![CDATA[d]]></KEY></LETTER>
<LETTER><TIME>633619249320396434</TIME><KEY><![CDATA[o]]></KEY></LETTER>
<LETTER><TIME>633619249352586200</TIME><KEY><![CDATA[w]]></KEY></LETTER>
<LETTER><TIME>633619249398995717</TIME><KEY><![CDATA[m]]></KEY></LETTER>
<LETTER><TIME>633619249411965380</TIME><KEY><![CDATA[<]]></KEY></LETTER>
<LETTER><TIME>633619249469781950</TIME><KEY><![CDATA[n]]></KEY></LETTER>
<LETTER><TIME>633619249489627097</TIME><KEY><![CDATA[ ]]></KEY></LETTER>
<LETTER><TIME>633619249602916322</TIME><KEY><![CDATA[b]]></KEY></LETTER>
<LETTER><TIME>633619249649638361</TIME><KEY><![CDATA[v]]></KEY></LETTER>
<LETTER><TIME>633619249666670810</TIME><KEY><![CDATA[<]]></KEY></LETTER>
<LETTER><TIME>633619249674952643</TIME><KEY><![CDATA[y]]></KEY></LETTER>
<LETTER><TIME>633619249711673978</TIME><KEY><![CDATA[ ]]></KEY></LETTER>
<LETTER><TIME>633619249715424242</TIME><KEY><![CDATA[ ]]></KEY></LETTER>
</IS>
```


Figure 28: An Example Output of the TypingCollector from which Typing Errors Must be Detected and Categorised

Manual classifications are also prone to subjective judgement - two people could categorise the same typing error as being different. In particular, when the error falls into two or more error types, the ambiguity is often solved arbitrarily, without any formal procedure of disambiguation (Logan, 1999). This lack of formal disambiguation methods also means that the data is open to a researcher's bias towards a particular error type. Without using a formal procedure to disambiguate this error to its true type, any findings the researcher may conclude will be based on a false sample of errors.

For large-scale studies to be carried out with accurate detection and classification of typing errors, it is necessary for the ExpECT method to be automated.

9.2.1 Current Method in Automated Categorisation of Typing Errors

Wobbrock and Myers' (2006) StreamAnalyzer made substantial advances in the automated analysis of character-level errors in input streams of unconstrained text entry evaluation. Their algorithm provided the groundwork on how to detect typing errors buried in the input stream by aligning and comparing letters from PT, IS and TT. Based on the MSD algorithm defined in (Soukoreff and MacKenzie, 2001), their StreamAnalyser offered the first program that took in PT and IS. The algorithm classified each letter in IS into Omission, Insertion, Substitution and No-Error and stated whether it was Corrected (subsequently deleted) or Uncorrected.

Section 8.3 used Wobbrock and Myers' StreamAnalyzer (Wobbrock and Myers, 2006) extensively to categorise typing errors made by 112 Children (1030 phrases, 2312 errors). Although it left 203 errors unclassified, this was due to the constraints given in their data collecting software TextTest (see Section 2.3.5.2) and not a shortfall of the analyser itself. The StreamAnalyzer found all errors within its own constraints, in a consistent and accurate manner.

Crucially, the analysis process was considerably faster than manually combing through the data. Even though the author had to enter one pair of PT and IS at a time into the StreamAnalyser and manually search and list the errors found, the process only took five days (compared to three weeks by manual classification). The StreamAnalyser does not perform disambiguation tasks when a PT/IS pair results in more than one possible answer. It produces all possible combinations of errors, and disambiguation was left to the user to resolve arbitrarily.

Wobbrock and Myers' (2006) work is an ideal starting point in developing a new automated categorisation program for the full-size keyboard. It is thorough, accurate and considerably faster than manual classification.

9.2.2 Requirement of the Analyser

Although Wobbrock and Myers' (2006) StreamAnalyzer and the new analyser both have the same aim of automatically classifying text input errors, there are also differences in their objectives. The StreamAnalyzer's objective was to detect and classify the three basic letter-level text input errors - Insertion, Omission and Substitution – independent of the input device. The new TypingAnalyser objective is to detect and classify all letter and word level typing errors (defined by ExpECT) for the full size QWERTY keyboard. These differences inevitably creates new requirements for the analyser as follows:

- *Detect and categorise typing errors*

  The analyser must be able to detect all typing errors from the Input Stream. It must be capable of categorising all the character and word-level error types defined by ExpECT.

- *Reduce error inflation and deflation*

  The analyser must prevent cases where a complex error is broken down into several smaller errors (e.g. Transposition Error being classified as an Inserted Letter and an Omitted Letter). It must also prevent the opposite cases where several smaller errors are erroneously classified as one complex error.

- *Use timing of keys pressed as part of the classification*

  The categorisation method ExpECT describes two error types (NT-Mu and CT-Mu) that require two keys to be pressed simultaneously. Therefore, the analyser must read in and use the timestamps to be able to distinguish these errors from other error types.

- *Use the location of the key on the keyboard in the classification*

  StreamAnalyzer was designed to be device independent. This meant that regardless of whether the data is from a full-size keyboard or a 12-key phone keypad, it is able to carry out the analysis on which letters were omitted, inserted or substituted. In contrast, the categorisation method ExpECT describes four error types that require the keys involved to be either next-to or close-to each other. Therefore, the TypingAnalyser must be able to look up a list of neighbouring letters for each key pressed.

- *Extend classifiable keys from letters only to function keys*

  TextTest (Wobbrock and Myers, 2006) - the data collector used by Wobbrock and Myers - disabled all keys except for the letters and the Enter key. Additionally, the pressing of the Enter key was only used to move onto the next phrase to copy. It was never added as a key press in

IS. Therefore, the StreamAnalyzer does not accept any function keys as an input. However, a substantial number of typing errors involving functional keys such as Enter and Shift have been noted in previous studies described in this thesis. Rather than disabling these keys altogether on the data collector, the analyser must be able to treat these function keys just like a letter key.

- *Perform disambiguation on the categorised errors*

  If a typing error is ambiguous, the TypingAnalyser must attempt to solve this based on predefined rules. When the TypingAnalyser fails to narrow the result down to one answer, a warning is given to the user.

- *Warn of likely misclassification or nonsense input*

  When the input is either vastly different from PT or contains many typing errors in a row it is likely that a nonsense text was entered by the participant. The TypingAnalyser should give warnings of these cases to the user.

- *Reduce the workload of manually checking the output by the TypingAnalyser*

  The TypingAnalyser must format the result in a way that makes it easier to spot typing errors. This consists of features such as not showing the list of all the keys classified if there were no errors, and indicating the number of errors found in each phrase typed.

- *Tabulate the result of categorised typing errors*

  A summary file must automatically produce reports on the following:

  - List of all errors found and their relevant information such as the intended and actual letter typed.
  - List of all phrases typed with relevant information such as the number of errors found per phrase.
  - List of participants with relevant information such as the number of errors each person made, etc.

The TypingAnalyser is a Java program that parses XML files of raw typing data produced by the TypingCollector (Chapter 6). It applies the categorisation method ExpECT (Chapter 8) to categorise all errors to one of 28 letter and word-level error types. The TypingAnalyser is capable of detecting and categorising not only character-level error types, but also word-level error types and character-level error types that consist of more than one character, such as Migration Error.

The TypingAnalyser also applies disambiguation rules defined in Section 9.8 to resolve as many ambiguities as possible. It gives warnings of those cases that may have been misclassified or require further attention from the user to check the results. For completeness, it also provides several aggregated error rates defined in (Soukoreff and MacKenzie, 2003).

The DataSummariser, also written in Java, parses the xml file created by the TypingAnalyser containing the categorised errors. The DataSummariser requires each case (a pair of PT and IS) to only have one set of results and gives warnings if this is not the case. It produces three comma-seperated-value (CSV) files:

1. errors.csv - a list of typing errors found
2. phrases.csv - a list of all typed phrases
3. participants.csv - a list of all participants analysed

The CSV files can be imported into programs such as SPSS (Statistical Package for the Social Sciences) and Microsoft™ Excel for further analysis.

Table 41 below summarises the features in all three typing analysis programmes. The TypingAnalyser is not only able to classify typing errors in much finer detail, but also makes significant steps to reduce the amount of work that is required to tabulated data.

Table 41: Summary of Features of the Three Automated Typing Analysers

Features	Soukoreff and MacKenzie (2003)	Wobbrock and Myers (2006)	TypingAnalyser
Calculate MSD	✔	✔	✔
Calculate aggregated error rates	✔	✔	✔
Character-level error analysis		✔	✔
Word-level error analysis			✔

Use of key press timings for classification			✔
Use of key location on the keyboard for classification			✔
Analysis of function key presses			✔
Carry out disambiguation			✔
Reduction of redundant results			✔
Warn of likely misclassifications			✔
Summarise data			✔

9.3 READING IN THE INPUT STREAM

As with Wobbrock and Myers' StreamAnalyzer (2006), for each case presented, the TypingAnalyser reads in the PT and IS from raw data. Crucially though, the TypingAnalyser deals with IS in a different manner to that of the StreamAnalyzer in three ways.

9.3.1 Text Stream as a List of Letters

Previous analysis of automated text input read each of the text streams as a string. Figure 29 below shows how the previous StreamAnalyzer read in the example given:

```
PT:                             thank you
IS (what was typed):            ALT_RIGHTty<hank yyou
IS(read in by StreamAnalyzer):ty<hank yyou
```

Figure 29: An Example PT and IS and How it is Read in by Wobbrock and Myers' StreamAnalyzer

This meant that it was not possible for the analyser to consider keys that represented functions, which have names longer than one character. Additionally, constraints must be added to the collection of typing data so that either the function keys are disabled altogether, or were not recorded by the data collecting software. Finally, the analyser was unable to examine values associated with each keystroke, such as the timing of each keystroke and what its neighbouring keys were.

In the previous chapter (Chapter 8), it was shown that 42 errors (Enter Error = 28 cases, Inserted Function = 24 cases) occurred that involved the function keys adjacent to the alphabetic keys on the keyboard. This is not an insignificant number of occurrences and therefore function keys were considered an important part of typing error analysis. The new TypingAnalyser accepts functional keys as part of IS.

The TypingAnalyser considers text streams as a list of letters. In particular the input stream is a long list of letters, where a letter could be an alphabetic character, a space, a symbol, a number or a function key (Table 42). Each letter also contains a list of its NT and CT letters and various other attributes required later in the analysis.

Table 42: Extract of IS in the New IS Format

IS[x]	Letter	NT Letters	CT Letters	Zero Time
0	ALT_RIGHT	SPACE, WIN_RIGHT	., ,, /	0
1	t	r, y	5, 6, f, g	0
2	y	t, u	6, 7, g, h	1
3	<	=,	], ENTER	0
4	t	r, y	5, 6, f, g	0
5	h	g, j	y, u, b, n	0

9.3.2 Flag the Input Stream for Simultaneous Key Presses

By definition, NT-Mu and CT-Mu depends on the intended letter and inserted letter to have been pressed at the same time. Since the TypingCollector's timestamp measures to the nearest 0.001 seconds, it is reasonable to assume that when two key presses have the same timestamp, they were pressed simultaneously. In addition, if the two keys were next to each other, it is reasonable to assume that the two keys were pressed simultaneously by one finger.

The TypingAnalyser passes forward through IS and flag any key that have the same timestamp as the previous letter (only the second key in the pair of keys is flagged regardless of which key was intended and which was the inserted letter). Figure 30 shows IS_1 where the first occurrence of `y` is flagged since its timestamp was the same as that of its preceding letter `t`.

```
→          0       010000000000000
IS1:   {ALT_RIGHT}ty<<than,k yyou
```

Figure 30: Letters hat have the Same Timestamp as the Previous Letter (Shown in Bold) are Flagged

9.3.3 Add Keyboard Location Information to Each Letter

To categorise errors involving the NT and CT letters (see Section 8.5.2.3 for definition), the NT and CT letters must be assigned to each key press in IS. A list of all keys and their relevant NT and CT letters specific to the keyboard layout used is required. A two-dimensional string array (called KEYS) of n x 3 size is used, where n is the number of keys on the keyboard, KEYS[n][0] holds the name of the key, KEYS[n][1] contains the list of corresponding NT letters and KEYS[n][2] contains a list of corresponding CT Letters. Table 43 shows an excerpt from one that is specific to a full-size QWERTY keyboard with a British/Irish key layout (see Appendix 1 for a picture of this layout).

Table 43: A Two-Dimensional Array Containing a List of NT and CT Letters for All Keys on a Keyboard

n	KEYS[n][0]	KEYS[n][1] (NT letters)	KEYS[n][2] (CT letters)
0	a	CAP_LOCK, s	q, w, \, z
1	b	v, n	g, h, SPACE
...	...	...	...

The TypingAnalyser then looks up this table and assigns the NT and CT letters to each key press in IS. This allows for location-dependent error categorisation to be done later in the process.

9.3.4 From Streams to Aligned Triplets

As with Wobbrock and Myers' (2006) analyser, for each pair of PT and IS, the TT is calculated by the TypingAnalyser. The Minimum String Distance (MSD - Section 3.2.1) is then calculated between the pair of PT and TT (Soukoreff and MacKenzie, 2001). Using Wobbrock and Myers' (2006) algorithms, all possible alignments between PT and TT to make the two strings the same (called 'Aligned Pairs') are computed. IS is then added to each of the Aligned Pairs to make 'Aligned Triplets'. Figure 31 below shows the differences between PT, IS, TT (a), Aligned Pair (b) and Aligned Triplet (c). Readers are directed to the original papers (Soukoreff and MacKenzie, 2001; Wobbrock and Myers, 2006) for further information on these computations.

```
PT: this winter
IS: ty<his wwintr<er
TT: this wwinter
```

a: PT, IS and TT at the Start

```
PT: this w_inter
TT: this wwinter
```

b: An Aligned Pair

```
PT: t__his w_int__er
TT: t__his wwint__er
IS: ty<his wwintr<er
```

c: An Aligned Triplet

Figure 31: An Example PT, IS, TT, and Their Aligned Pair and Aligned Triplet

One pair of PT and TT can have anything from one to several Aligned Pairs. For each Aligned Pair created, one Aligned Triplet is computed. It is entirely possible for a PT/TT pair to have several hundreds of Aligned Pairs. For simplicity, the TypingAnalyser terminates the computation of Aligned Pairs after computing five Aligned Pairs. Figure 32 shows an example where three Aligned Pairs were computed for PT and TT, which in turn resulted in three Aligned Triplets. The Categorisation methods described in Section 9.4 to Section 9.9 are then performed for each Aligned Triplet.

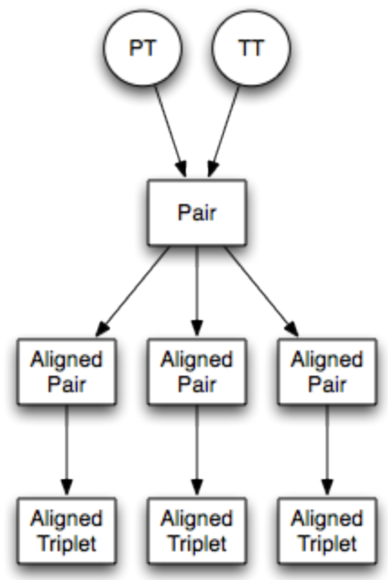


Figure 32: One Pair of PT and TT can have Several Aligned Pairs - Each AlignedPair Only has One AlignedTriplet (IS has been Removed for Clarity)

9.4 CLASSIFYING LETTER-LEVEL ERRORS - DETERMINEERRORS()

The TypingAnalyser now calls on the method `determineErrors()` to detect and categorise character-level errors. The method is based upon the mechanism defined by Wobbrock and Myers (Wobbrock and Myers, 2006) in detecting Omission, Substitution and Insertion errors, but classifies errors in much finer detail. For each AlignedTriplet the TypingAnalyser compares PT and IS letter by letter, calculating the intended PT letter and what was actually typed in IS. As each error (or NoError) is classified, it is added to resultList (a list of all the classifications made already for that AlignedTriplet). Table 44 below shows how Wobbrock and Myers' original three error types have been divided in this stage of categorisation.

Table 44: Wobbrock and Myers' Original Three Error Types are Divided into Finer Detail in TypingAnalyser (Error Types Discussed Further in this Chapter are Shown in Bold)

Error types categorised by Wobbrock and Myers (2006)	Errors categorised by TypingAnalyser at the end of determineErrors()
Omission	Omitted Letter Omitted Space
Insertion	Inserted Letter Inserted Symbol Inserted Function Inserted Space Enter Error **NT-Mu and CT-Mu** **Duplicated Letter and Duplicated Space** **Execution Error**
Substitution	Substituted for a Letter Substituted for a Symbol Substituted for a Space Substituted for a Function Capitalisation Error **NT-S and CT-S** **Doubling Error** **Transposition Error**

Since this method is based on that of Wobbrock and Myers' (2006) it is also able to determine whether each error was corrected or not. The term 'corrected' does not refer to the typing error being fixed, but only that the error was subsequently deleted in an attempt to fix the error.

The `determineErrors()` method is able to categorise errors into all the singular character-level error types. These are carried out by comparing the character found in IS with either the intended letter in PT or the previous letter in IS. When these singular character-level errors are detected, they are determined as an Omission, an Insertion or a Substitution. They are then tested for various error types within its relative error group (those shown in bold in Table 44).

9.4.1 Determine NT-Mu/CT-Mu (Insertion)

When an Insertion is discovered, it is first tested to see if the error can be an NT-Mu or CT-Mu (see Section 8.5.2.3 for definitions). The intended letter and the extra letter typed must be either Next-To or Close-To each other on the keyboard. The two keys must also have the same timestamp. The TypingAnalyser uses the ZeroTime flag on the IS (Section 9.3.2) to determine this. If the Insertion is not found to be NT-Mu or CT-Mu, it is then tested for Duplicated Letter.

9.4.2 Detecting Duplicated Letter/Space (Insertion)

Duplicated Letter Error occurs when an intended letter is produced twice, such as when '`shoe`' is typed as '`sshoe`'. Wobbrock and Myers' (2006) algorithm classified this as an Insertion. It is also unknown whether it was the first '`s`' or the second '`s`' that was inserted, so their StreamAnalyzer produces two possible results to represent both possibilities. However, the ExpECT method defines these cases specifically as Duplicated Letter or Duplicated Space. The method assumes that, when such error occurs, the first key press was intended correctly, and that it is the second key press that was erroneous. As such, the TypingAnalyser computes only one resultList.

9.4.3 Detecting Execution Errors (Insertion)

An Execution Error occurs when a key is held down for too long, resulting in multiple letters being produced from that key. In TypingAnalyser, any key press of the same letter that occurs three or more times in a row is classified as Execution Error. This is detected by scanning each resultList for a continuous sequence of the same key. An exception is that of holding down the Caps lock or the Shift key while trying to produce an uppercase letter, which is common and necessary. If either of these keys are held down to produce an intended uppercase letter, it is classified as NoError. If, on the other hand, the keys were held down to produce an uppercase letter when a lowercase letter was required, this is classified as an Inserted Function Error.

9.4.4 Determine NT-S/CT-S (Substitution)

When a substitution is discovered, the intended key and the actual key pressed are compared. If they are found to be either Next-To or Close-To each other on the keyboard, then the classification of the error changes from Substituted Letter to 'NT-S' or 'CT-S'.

9.4.5 Doubling Error (Substitution)

A Doubling Error (where the wrong letter is doubled - Section 8.5.2.7) can occur either before or after the letter originally intended to be doubled as Figure 33 shows.

```
Case 1:                    Case 2:
PT: this book is           PT: this book is
IS: this bbok is           IS: this bokk is
```

Figure 33: Doubling Error Can Occur Either Before or After the Letters Originally Intended to be Doubled

Previously, these would have been classified as Substitutions and buried under many other Substitutions. Instead, `determineErrors()` checks to see if the following two conditions are fulfilled:

1. *Current letter in IS == previous letter in IS?*
2. *Is the intended letter in PT flagged as double letter?*

When these two conditions are met, it is classified as Doubling Error. Otherwise, the error is passed on to the `determineSubstitution()` method to be classified into another Substitution error type.

9.4.6 Transposition Error (Substitution)

Transposition Error (where two consecutive letters are swapped - see Section 8.5.2.2) previously suffered from error inflation and ambiguities. Figure 34 shows the result of Wobbrock and Myers' analysis for the example PT ('`team`') and IS ('`taem`').

```
PT: team
IS: taem
```

resultList1:	resultList2:	resultList3:
U No Error(t,t)	U No Error(t,t)	U No Error(t,t)
U Substituted Letter(e,a)	**U Inserted Letter(-,a)**	**U Omitted Letter(e,-)**
U Substituted Letter(a,e)	U No Error(e,e)	U No Error(a,a)
U No Error(m,m)	**U Omitted Letter(a,-)**	**U Inserted Letter(-,e)**
	U No Error(m,m)	U No Error(m,m)

Figure 34: The Three resultLists of 'team->taem' as Classified by StreamAnalyzer. Errors are Shown in Bold (U = Uncorrected)

Instead, when the TypingAnalyser detects a Substitution, it checks to see if the following two conditions are met:

1. *Current letter in IS == next letter in PT*
2. *Next letter in IS == current letter in PT*

Figure 35 below shows an example of PT and IS where these two conditions are met.

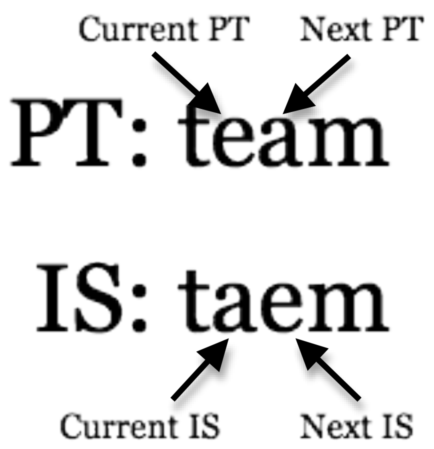


Figure 35: Conditions for a Transposition Error.

When these conditions are met, the error is classified as Transposition Error. It should be noted that the three triplets are still created by the TypingAnalyser as shown in the previous Figure 34. The TypingAnalyser only changes the result of the first triplet (resultList1) to include a Transposition Error. The two other triplets remain the same as those calculated by Wobbrock and Myers' as shown in Figure 36. However, resultList2 and resultList3 will be deleted later in `deleteTEDuplicates()` (Section 9.8.2).

```
                         PT: team
                         IS: taem

resultList1:             resultList2:             resultList3:
U No Error(t,t)          U No Error(t,t)          U No Error(t,t)
U Transposition(ea,ae)   U Inserted Letter(-,a)   U Omitted Letter(e,-)
U No Error(m,m)          U No Error(e,e)          U No Error(a,a)
                         U Omitted Letter(a,-)    U Inserted Letter(-,e)
                         U No Error(m,m)          U No Error(m,m)
```

Figure 36: The Three resultLists After Transposition Error is Classified in resultList1

9.4.7 Not Classifying Shift and Caps Lock as Errors When Intended

All function key presses are classified as either Inserted Function or Substituted for a Function, with one exception. When an uppercase letter is intended (such as 'I') if the TypingAnalyser encounters either Shift or Caps lock key press, these are not classified as errors. The user is required to press Shift or Caps lock to produce the necessary uppercase letters. This rule does not apply when the intended letter is in lowercase. For this, the Shift or Caps lock key presses will be classified as an error.

9.4.8 Determining Errors by the Letters Involved

If an error type does not fit into the error types already mentioned, it is passed on to either `determineOmission()`, `determineInsertion()` or `determineSubstitution()` to be further categorised according to the error involved.

`determineOmission()` only separates incoming errors into either letters or spaces, since the PT only contains letters and spaces. For `determineInsertion()` and `determineSubstitution()`, where there are no restrictions as to what could be entered, the methods must split the character further (letter, space, function, symbol, etc.).

9.4.9 At the End of determineError()

As each letter in IS is being categorised into an error type or NoError, they are added to resultList. Once `determineError()` has gone through an entire Aligned Triplet, each Aligned Triplet's resultList will look similar to that shown in Figure 37:

```
                              resultList1:
                              Uncorrected Inserted Function(-,ALT_RIGHT)
                              Corrected NT-Mu(t,ty)
                              Uncorrected No Error(h,h)
                              Uncorrected No Error(a,a)
PT: thank you                 Uncorrected No Error(n,n)
IS: ALT_RIGHTty<hank yyou     Uncorrected No Error(k,k)
                              Uncorrected No Error( , )
                              Uncorrected No Error(y,y)
                              Uncorrected Duplicated Letter(y,yy)
                              Uncorrected No Error(o,o)
                              Uncorrected No Error(u,u)
```

Figure 37: Example of resultList After determineErrors() is Completed

9.5 CALCULATION OF ERROR RATES

After all Aligned Triplets are categorised, various aggregated error rates are calculated. These calculations are done at this stage since aggregated error rates only require the number of Correct letters (C), Incorrect and Fixed letters (IF), Incorrect and Not Fixed letters (INF) and Fixes (F). The aggregated error rates calculated by the TypingAnalyser (by `setErrorStats()` in triplet.java) are listed in Table 45 below together with their formulae and definitions. Readers are directed to (Soukoreff and MacKenzie, 2003) for further information on these error rates.

Table 45: Aggregated Error Rates Calculated by the TypingAnalyser

Term	Name	Definition	Formula
C	Corrected letters	Number of No Errors	
F	Fixes	Number of '<' (Backspace)	
IF	Incorrect and Fixed	Number of corrected errors	
INF	Incorrect and Not Fixed	Number of uncorrected errors	
TER	Total Error Rate	Ratio of errors to the total number of keys pressed	(INF+IF)/(C+INF+IF)

CER	Corrected Error Rate	Ratio of errors corrected to the total number of keys pressed	INF/(C+INF+IF)
UER	Uncorrected Error Rate	Ratio of errors not corrected to the total number of keys pressed	IF/(C+INF+IF)
CE	Correction Efficiency	Ease at which the participant performed error corrections	IF/F
PC	Participant Conscientiousness	Ratio of corrected errors to the total number of errors	IF/(IF+INF)
UB	Utilised Bandwidth	Proportion of bandwidth representing useful information transfer	C/(C+INF+IF+IF)
WB	Wasted Bandwidth	Proportion of bandwidth representing wasted information transfer	(INF+IF+F)/(C+INF+IF+F)

9.6 CLASSIFYING SPLIT CHARACTER-LEVEL ERRORS

Some character-level errors appear as two separated errors in resultList, which after `determineErrors()`, remain misclassified. These errors are Migration, Interchange and Alternating Errors. For example, a Migration Error, where one letter moves across IS, is classified as an Omitted Letter, then an Inserted Letter, as Figure 38 shows:

```
PT: hospital
IS: hosptail
```

```
resultList1:
Uncorrected No Error(h,h)
Uncorrected No Error(o,o)
Uncorrected No Error(s,s)
Uncorrected No Error(p,p)
Uncorrected Omitted Letter(i,-)
Uncorrected No Error(t,t)
Uncorrected No Error(a,a)
Uncorrected Inserted Letter(-,i)
Uncorrected No Error(l,l)
```

Figure 38: resultList for a Migration Contains Two Errors After determineError()

Wobbrock and Myers' StreamAnalyzer ends its analysis at this point. An error that is actually one Migration Error is therefore counted as two character-level errors, creating error inflation.

In contrast, the new TypingAnalyser takes the analysis further by searching for these character-level errors that are separated by more than one NoError. The

TypingAnalyser makes passes through each resultList for characteristics associated with Migration, Interchange and Alternating Error.

9.6.1 Detection of Migration

Figure 39 below shows the two patterns of Migration. One is for the migrant letter to appear later than intended (to the right of the intended position) (case 1 in Figure 39). The other is for the migrant letter to appear in IS earlier than intended (to the left of the intended position), shown here as case 2 in Figure 39. As the resultList for each case shows, these are classified by `determineError()` as either an Omission and then Insertion of the migrant letter, or and Insertion then an Omission of the migrant letter.

```
Case 1:     PT: hospital                 Case 2:     PT: hospital
            IS: hosptail                             IS: hiosptal

resultList1:                             resultList1:
Uncorrected No Error(h,h)                Uncorrected No Error(h,h)
Uncorrected No Error(o,o)                Uncorrected Inserted Letter(-,i)
Uncorrected No Error(s,s)                Uncorrected No Error(o,o)
Uncorrected No Error(p,p)                Uncorrected No Error(s,s)
Uncorrected Omitted Letter(i,-)          Uncorrected No Error(p,p)
Uncorrected No Error(t,t)                Uncorrected Omitted Letter(i,-)
Uncorrected No Error(a,a)                Uncorrected No Error(t,t)
Uncorrected Inserted Letter(-,i)         Uncorrected No Error(a,a)
Uncorrected No Error(l,l)                Uncorrected No Error(l,l)
```

Figure 39: Two Patterns of Migration after determineError()

`SearchForMigration()` in resultList.java, searches for either an Omitted Letter/Space or an Inserted Letter/Space. When one is found, the method skips the next letter (by definition, the intended position and the actual position must be at least one letter apart - see Section 8.5.2.9). The method then scans for either an insertion (of the same letter) if an omission was first found, or an omission of the same letter if insertion was first found.

The algorithm for searching for Migration imposes two restrictions in searching for the second error:

1. *The search must terminate if '<' (Backspace) is encountered in IS*

   The entire action of an error must be carried out in one process - i.e. both the insertion and omission must occur in one smooth process with no interruptions. A Backspace indicates a break in this process.

2. *The search is limited to the word in which the first error was found, through to (and including) the next word, but no further*

Once the matching error is found, the method reclassifies these errors as 'Migration Error across X letters'. This error replaces the first error in found in resultList, and the second error is deleted.

9.6.2 Detecting Interchange

Similar to Migration, Interchange Error consists of two character-level errors that are at least one letter apart. An Interchange Error occurs when two letters are swapped around across a minimum of one letter. Figure 40 below shows an example of an Interchange Error.

```
                                      resultList1:
                                      Uncorrected No Error(t,t)
PT: through                           Uncorrected No Error(h,h)
IS: thgourh                           Uncorrected Substituted Letter(r,g)
                                      Uncorrected No Error(o,o)
                                      Uncorrected No Error(u,u)
                                      Uncorrected Substituted Letter(g,r)
                                      Uncorrected No Error(h,h)
```

Figure 40: Example of Interchange Across Two Letters

Interchanges are classified by `determineErrors()` as two substitutions. Therefore, the method `searchForInterchange()` starts by forward searching the resultList for a substitution involving letters or spaces. Symbols, numbers and functions are not considered since these do not appear in PT.

Once a substitution is found, the method skips the next letter and then continues along resultList for the second substitution. This second substitution must involve the same letters as the first substitution. The method `searchForInterchange()` imposes the same restrictions as `searchForMigration()` does in searching for the second error.

1. *The search must terminate if '<' (Backspace) is encountered in IS*

   The entire action of an error must be carried out in one process - i.e. both the insertion and omission must occur in one smooth process with no interruptions. A Backspace indicates a break in this process.

2. *The search is limited to the word in which the first error was found, through to (and including) the next word, but no further*

Once the matching Substitution is found, the method reclassifies the first error as Interchange across X letters. The second Substitution is removed from the resultList.

9.6.3 Detecting Alternating Error

An Alternating Error is defined as when a string sequence of '`xyx`' is typed '`yxy`' instead. It can only occur in words that have the string sequence '`xyx`' such as the word '`there`'. When '`there`' is typed as '`threr`', the method `alignPair()` (in AlignedPairs.java) will compute two possible alignments between PT and TT. This results in two resultLists, as shown in Figure 41 below.

```
                              PT: there
                              IS: threr
resultList1:                        resultList2:
Uncorrected No Error(t,t)           Uncorrected No Error(t,t)
Uncorrected No Error(h,h)           Uncorrected No Error(h,h)
Uncorrected Inserted Letter(-,r)    Uncorrected Omitted Letter(-,e)
Uncorrected No Error(e,e)           Uncorrected No Error(r,r)
Uncorrected No Error(r,r)           Uncorrected No Error(e,e)
Uncorrected Omitted Letter(e,-)     Uncorrected Inserted Letter(-,r)
```

Figure 41: Two resultLists for 'there->threr' After determineErrors()

The method `searchForAlternation()` (in ResultList.java) will therefore only search for an Alternating Error if PT contains the string sequence 'xyx'. It then searches for either an [Inserted Letter, No Error, No Error, Omitted Letter] sequence, or [Omitted Letter, No Error, No Error, Inserted Letter] sequence in resultList.

If such a combination is found the letters involved in the four-error sequence are analysed. The letter of the first error must equal the letter of the third error, and the letter of the second error must equal that of the fourth error, as shown in Table 46 below.

Table 46: The Letters of the Four-error Sequence Found by searchForAlternation() that Matches the Given Conditions

	Error Type	FROM	TO
error 1	Uncorrected Inserted Letter	-	r
error 2	Uncorrected No Error	e	e
error 3	Uncorrected No Error	r	r
error 4	Uncorrected Omitted Letter	e	-

When all three conditions are met, the error is classified as 'Alternating Error(xyx, yxy)'. This new classification replaces all four of the four-error sequence. The resultLists from Figure 41 after this classification are shown in Figure 42 below:

```
PT: there
IS: threr

resultList1:                         resultList2:
Uncorrected No Error(t,t)            Uncorrected No Error(t,t)
Uncorrected No Error(h,h)            Uncorrected No Error(h,h)
U Alternating Error(ere,rer)         U Alternating Error(ere,rer)
```

Figure 42: resultLists for the Example 'there -> threr' After the Alternation Classification

Figure 42 above shows that both resultLists now have exactly the same list of errors. The TypingAnalyser will reduce these duplications and redundant resultLists later in Section 9.8.2.

9.7 DETECT AND CLASSIFY WORD-LEVEL ERRORS

At the end of all character-level errors analysis, word-level errors are classified as several character-level errors (Figure 43). If no further analysis were carried out (as in the case of StreamAnalyzer), one word-level error, such as a Substituted Word, would be logged as several Substituted Letters.

```
                           resultList1:
                           Uncorrected No Error(t,t)
                           Uncorrected No Error(h,h)
                           Uncorrected No Error(e,e)
PT: the book was           Uncorrected No Error( , )
IS: the shoe was           Uncorrected Substituted Letter(b,s)
```

```
Uncorrected Substituted Letter(o,h)
Uncorrected No Error(o,o)
Uncorrected Substituted Letter(k,e)
Uncorrected No Error( , )
Uncorrected No Error(w,w)
Uncorrected No Error(a,a)
Uncorrected No Error(s,s)
```

Figure 43: resultList Containing a Word-level Error After All Character-level Error Analysis are Complete (Note how One Word-level Error is Classified as Several Character-level Errors Still)

Not only does this make the word-level error harder to find among hundreds of other Substitutions, it also vastly inflates the total number of errors found. Since this is counter-productive in trying to accurately count the number of errors created by each participant, the TypingAnalyser's next contribution to categorisation of error types is to pass through each resultList again, this time to detect and categorise word-level errors.

9.7.1 What is a Word

Before one can search for a word, the definition of what a word is must clearly be defined. For a string of characters to be understood as a word, it must first consist of at least two characters (with the exception of 'a' and 'I'). Additionally, a the word must be found in one of two lists:

1. A given dictionary - the TypingAnalyser can use any given text file that has one word per line. Currently, the TypingAnalyser uses a dictionary (dictionary.txt) that contain the Oxford dictionary's '3000 most important words' (Hornby, 2007) and some additional words (real words typed by participants) - see Appendix 4 for the list of words the file contains. This file can be edited to add more words when necessary.
2. Words shown in PT already - for each run of analysis, the TypingAnalyser keeps a list of all the words that appear in PT that it reads in. It is assumed that all PT will only contain complete words by design. This list is only kept for the duration of one participant's analysis and is deleted before analysing the second participant.

If a match for a given string is not found in either list, it is considered not to be a word and therefore discarded from the search for a word-level error.

9.7.2 In Search of Potential Word-Level Errors (analyseWordLevelErrors() in ResultList.java)

The first step in searching for a word-level error is to scan through each resultList for a sequence of errors that are potentially word-level errors. The TypingAnalyser proceeds through resultList (Figure 36) in search of consecutive errors. When more than one error is found in a row, an errorUnit (shown in bold) is created, which contains the entire sequence of continuous errors (Figure 44 below):

```
                                resultList1:
                                Uncorrected No Error(t,t)
                                Uncorrected No Error(h,h)
                                Uncorrected No Error(e,e)
                                Uncorrected No Error( , )
PT: the book was                Uncorrected Substituted Letter(b,s)
IS: the shoe was                Uncorrected Substituted Letter(o,h)
                                Uncorrected No Error(o,o)
                                Uncorrected Substituted Letter(k,e)
                                Uncorrected No Error( , )
                                Uncorrected No Error(w,w)
                                Uncorrected No Error(a,a)
                                Uncorrected No Error(s,s)
```

Figure 44: ResultList for the Given Example, Indicating the Initial errorUnit in Bold

On some occasions, one Substituted Letter could appear to be a Substituted Word, as in '`book`' is typed as '`cook`' or '`hook`'. This would result in an ambiguity between a Substituted Letter and a Substituted Word. For the purpose of disambiguation between these two error types, the TypingAnalyser classifies such errors as a Substituted Letter. Two or more letters must be altered for the classification of Substituted Word. This is why the search starts off with looking for two or more errors in a row.

9.7.3 Calculating the Intended and Actual Words

Next, the errorUnit is examined to compute the intended word (FROM) and the typed word (TO). For the example given in Figure 44 FROM is '`bo`' and TO is '`sh`'. These two strings are tested to see if either form real words.

9.7.4 Testing a String to See if it is Real and Complete Word

The string and TO and FROM are now parsed to `isWord()` in resultList.java. `isWord()` first searches through predefined list of words (found in oxford3000.txt) for a match. If no match is found in dictonary.txt a list of all the words so far shown in PT is compared for a match. If the string is found in either of the two lists, or the string is the word 'a' or 'I', then the method `isWord()` will return true.

If neither TO nor FROM is a real word, the TypingAnalyser makes a backward pass through the resultList to add more errors to the errorUnit. It will continue adding errors until it encounters a ' ' (space) or '-' in PT, or if the method reaches the start of resultList.

In the example shown in Figure 44 no error is added as the intended letter of the error immediately before the errorUnit is a space. Once the backward pass is complete, TO and FROM are once again computed and tested to see if either string is a real word (by `isWord()`).

If neither strings return true from `isWord()`, then the TypingAnalyser makes a forward pass through resultList from the end of errorUnit. Errors are added to errorUnit until a space is encountered or an errorUnit consisting of corrected errors encounters an uncorrected error. Figure 45 below shows the errorUnit after the forward passes have terminated on encountering the UncorrectedNoError(,).

```
                         resultList1:
                         Uncorrected No Error(t,t)
                         Uncorrected No Error(h,h)
                         Uncorrected No Error(e,e)
                         Uncorrected No Error( , )
PT: the book was         Uncorrected Substituted Letter(b,s)
IS: the shoe was         Uncorrected Substituted Letter(o,h)
                         Uncorrected No Error(o,o)
                         Uncorrected Substituted Letter(k,e)
                         Uncorrected No Error( , )
                         Uncorrected No Error(w,w)
                         Uncorrected No Error(a,a)
                         Uncorrected No Error(s,s)
```

Figure 45: ErrorUnit Shown in Bold - FROM is Now 'book' and TO is Now 'shoe', Which are Real Words

The strings TO and FROM are computed and checked by `isWord()` to see if they are real words again. As the example in Figure 45 shows, FROM is now `'book'` and TO is now `'shoe', which` are both real words. If neither string returns true at this point, it is assumed that the errors found were not of word-level and so the search for a word error is terminated, and its categorisations remain unedited.

However, if after any of these three passes through resultList, either TO and/or FROM is found to be a real word, the TypingAnalyser will test these strings further to see if they can be classified as either an Omitted, Substituted or Inserted Word.

9.7.5 Classifying Substituted Word

For a word-level error to be classified as a Substituted Word, firstly both TO and FROM must be real words. A word in PT cannot be substituted by a nonsense word. Secondly, TO and FROM must be different words. When these two conditions are met, it is assumed that the errorUnit contains a Substituted Word. The errorUnit in resultList is replaced by Substituted Word (FROM,TO).

The TypingAnalyser then attempts to establish where the TO string came from. According to the categorisation method ExpECT, TO can either come from the list of words in the current PT (This phrase - listed in `PTWords`), or from the PTs shown to the participants (from Another place - listed in `previousPRWords`) or from an Unknown source. The source is indicated by either T, A or U. For the example given of `boot->shoe`, the error will now be written as 'Substituted Word (U) (boot,shoe)' as the word 'shoe' is not found in either the current phrase or any phrases before.

9.7.6 Classifying Inserted Words

For an errorUnit to be classified as an Inserted Word the errorUnit must satisfy the following conditions:

1) *TO must be a real word*

AND

2) i) *TO must be a duplication of the previous word found in IS*

OR

ii) *The errorUnit must consist of insertion error types and NoErrors.*

Figure 46 below demonstrates these cases. In the first case (Case A) the errorUnit satisfies conditions 1) and 2i) and so the errorUnit is replaced by Uncorrected Inserted Word(-,one) and Uncorrected Inserted Space(-,). The source of the inserted word is then determined in the same manner as that described in classifying a Substituted Word (Section 8.6.2).

```
Case A:     PT: the book
            IS: the one book
errorUnitA:
Uncorrected Inserted Letter(-,o)
Uncorrected Inserted Letter(-,n)
Uncorrected Inserted Letter(-,e)
Uncorrected Inserted Space(-, )
```

```
Case B:     PT: the book
            IS: the the book
errorUnitB:
Uncorrected Inserted Letter(-,t)
Uncorrected Inserted Letter(-,h)
Uncorrected Inserted Letter(-,e)
Uncorrected Inserted Space(-, )
```

Figure 46: Partial resultLists of Two Examples of an Inserted Word After determineErrors()

In the second case (Case B) the errorUnit satisfies the conditions 1) and 2ii), since the word inserted is the same word as the word typed just before the errorUnit. This is classified as 'Uncorrected Duplicated Word(the , the the)'.

9.7.7 Classifying Omitted Word

For an errorUnit to be classified as an Omitted Word, it must satisfy the following conditions:

1) *FROM must be a real word*

AND

2) *FROM must be found in the current PT*

AND

3) *The errorUnit must only contain Omitted Letters, Omitted Space and NoErrors.*

If all three conditions are met, it is classified as Omitted Word(FROM,-). Figure 47 below shows an example PT and IS, with its resultList before and after this classification.

```
PT: the book was
IS: the was
```

BEFORE	AFTER
resultList1:	resultList1:
Uncorrected No Error(t,t)	Uncorrected No Error(t,t)
Uncorrected No Error(h,h)	Uncorrected No Error(h,h)
Uncorrected No Error(e,e)	Uncorrected No Error(e,e)
Uncorrected No Error(,)	Uncorrected No Error(,)
Uncorrected Omitted Letter(b,-)	Uncorrected Omitted Word(book,-)
Uncorrected Omitted Letter(o,-)	Uncorrected Omitted Letter(,-)
Uncorrected Omitted Letter(o,-)	Uncorrected No Error(w,w)
Uncorrected Omitted Letter(k,-)	Uncorrected No Error(a,a)
Uncorrected Omitted Letter(,-)	Uncorrected No Error(s,s)
Uncorrected No Error(w,w)	
Uncorrected No Error(a,a)	
Uncorrected No Error(s,s)	

Figure 47: resultList Before and After the errorUnit was Classified as an Omitted Word

Instead of five letter-level errors, the resultList now contains one word-level error which is much more accurate in the number of errors in resultList. It is also far more useful to a researcher to see that the participant omitted an entire word, than a sequence of omitted letters that may or may not have been an omitted word.

9.8 REDUCING DUPLICATIONS AND REDUNDANCIES

During the classification of typing errors, some pairs of PT and IS end up with multiple resultLists that are the same, such as the example given in Detecting Alternating Error (Section 8.5.2.6). Others may contain one resultList that has a word-level error, or other multi-character errors (such as Transposition Errors) and one or more resultLists that continued to have the multiple single-character errors. In both cases, these redundant resultLists must be deleted to reduce ambiguity.

One assumption that is made here, for the purpose of reducing ambiguities, is as follows:

If there is more than one possible combination of errors (resultList) for the same typing error, then the one that has the least number of errors is chosen.

This assumption is to reduce error inflation. If there are two resultLists for a pair of PT and IS, where one contain a multi-character error (such as Omitted Word), and the other contains several single-character-level errors that represent the same typing error, then the first resultList is always chosen.

9.8.1 Deleting Duplicated ResultLists

Deleting duplicated resultLists within the same pair of PT and IS is very simple. If an Aligned Triplet has more than one resultList, each resultList is compared in turn to all other resultLists for the pair. If an exact match occurs, the second matching resultList is deleted. For the example of Alternating Error found in Figure 42, the second resultList is deleted.

9.8.2 Deleting Redundant ResultLists of Transposition Errors

Figure 36 in Section 9.4.6 shows that there are three resultLists for this example of `'team'` being typed as `'taem'`. The first resultList contains the error Transposition Error(ea,ae) and is the most accurate. The other two resultLists break down the same error into two character-level errors. Since this causes error inflation the two latter resultLists should be deleted.

For each PT/IS pair, the method `deleteTEDuplicates()` (in AlignedTriplets.java) searches through all the resulLists belonging to that pair for a Transposition Error. It then searches through all other resultLists for that pair, for the error sequences matching one of the following conditions:

- Inserted Letter/Space - No Error - Omitted Letter/Space (resultList2)
- Omitted Letter/Space - No Error - Inserted Letter/Space (resultList3)
- Zero-time Inserted Letter - Omitted Letter
- Omitted Letter - Zero-time Inserted Letter

The letters involved in all the errors mentioned above must all match with the letters in the Transposition Error.

Once a match is found, the later resultLists are deemed to be redundant, and thus removed. In the example of Figure 36, both resultList2 and resultList3 match these conditions and so are deleted, leaving only resultList1 as the possible result of '`team`' to '`taem`'.

9.8.3 Deleting Redundant ResultList Word-Level Errors

Similar to Transposition Error redundancies, PT/IS pairs containing a word-level error may also have redundant resultLists. This usually occurs when the first letter (or first few letters) of the erroneous word and the word after match. Such cases only occur in either Omitted or Substituted Word. Figure 48 below shows an example of this case. For simplicity, the error types have been abbriviated (UNE = Uncorrected No Error, UOmSpace = Uncorrected Omitted Space, UOmWord = Uncorrected Omitted Word, UOmSpace = Uncorrected Omitted Space).

```
PT: in the tall
IS: in tall
```

resultList1:	resultList2:	resultList3:
UNE(i,i)	UNE(i,i)	UNE(i,i)
UNE(n,n)	UNE(n,n)	UNE(n,n)
UOmSpace(,-)	**UNE(,)**	**UNE(,)**
UOmWord(the,-)	**UOmWord(the,-)**	**UNE(t,t)**
UNE(,)	**UOmSpace(,-)**	**UOmLetter(h,-)**
UNE(t,t)	**UNE(t,t)**	**UOmLetter(e,-)**
UNE(a,a)	UNE(a,a)	**UOmSpace(,-)**
UNE(l,l)	UNE(l,l)	**UOmLetter(t,-)**
UNE(l,l)	UNE(l,l)	UNE(a,a)
		UNE(l,l)
		UNE(l,l)

Figure 48: resultLists of 'in the tall' to 'in the' After All Classification ('errorZone' is Indicated in Bold)

For each PT/IS pair, `deleteDuplicateWord()` (in AlignedTriplets.java) searches in each resultList for a Omitted or Substituted Word. If such an error is are found, the method then needs to search through all other resultLists for the pair to determine if this particular error is the only discrepancy between the resultLists. As Figure 48 shows, the Omitted or Substituted Words have varying total numbers of errors and appear in varying locations in the resultLists. Therefore, a simple comparison of a section from each resultList starting from the same location as the word error found will not suffice.

The method first assumes that the word error is the only error in all resultLists. It then computes an errorZone for each resultList. The errorZone is computed in a way that the errors before and after it will be the same for all resultLists if the assumption is correct.

In analysing resultList2 in Figure 48, it is found that the portion of resultList1 before the errorZone matches that of resultList2 before the errorZone. Additionally, the portion of resultList1 after the errorZone matches that of resultList2 after the errorZone. It is therefore, assumed that resultList2 is redundant, and is removed. The example in Figure 48 shows that resultList3 also satisfies the same conditions and therefore is deleted, leaving only resultList1 as the possible answer.

9.8.4 Reducing resultList with Error Inflation

The assumption 'if there is more than one possible combination of errors (resultList) for the same typing error, then the one that has the least number of errors is chosen' is applied further here to reduce ambiguities. For each Aligned Pair that contains a word-level error, the resultList with the least number of errors should be kept, and the other resultLists deleted.

For each PT/IS pair, `reduceWordTriplets()` (in AlignedTriplets.java) examines all resultLists for the pair, to compute the minimum number of errors found in any resultList. It then removes any resultLists that contain a greater number of errors than the minimum. It will however not delete any resultLists that contain the same number of errors as the minimum.

9.9 REDUCING THE WORKLOAD OF THE USER

Besides detecting and categorising typing errors, the aim in this chapter is to reduce the work by the user to convert raw data to tabulated data. In this section, two strategies implemented to reduce the user's workload are discussed.

9.9.1 Give Warning of Cases That May Require Checking

Although the TypingAnalyser is designed to be as accurate as possible in its classifications and disambiguations, there are cases where a human-check or a decision is required. In these cases, the TypingAnalyser will produce appropriate warnings in result.xml. The user can search through result.xml for the term 'WARNING' to double check on just these cases, rather than having to check every single case tested.

The TypingAnalyser gives three kinds of warnings:

- 'MORE THAN ONE RESULT LIST - PLEASE REDUCE TO ONE' - for each PT/IS pair, there must only be one resultList. Although the TypingAnalyser takes several steps to reduce these ambiguities, a few (such as two resultLists containing the same number of errors) remain ambiguous. In such a case, the TypingAnalyser gives a warning to the user that one or more resultList must be deleted.
- 'MORE THAN 50% TOTAL ERROR RATE - PLEASE CHECK' - the TypingAnalyser is unable to detect when a nonsense word is entered. It will still attempt to categorise the 'errors' between PT and this nonsense IS. However, it is likely that a researcher using the TypingAnalyser would wish to remove nonsense cases from their sample. Therefore, when a case has a high Total Error Rate (TER) a warning is given in result.xml.
- 'MORE THAN 4 ERRORS IN ROW - PLEASE CHECK' - four typing errors in a row occur relatively infrequently and their existence usually indicates that something has gone wrong. This could be that the TypingAnalyser did not pick up a word-level error, or that the participant entered nonsense words. In either case, the TypingAnalyser gives a warning so that the cause of such a high number of errors can be accurately determined by a manual inspection.

9.9.2 Tabulation of Errors Found

Once all typing errors are categorised and disambiguated as much as possible, a researcher can use the DataSummariser to tabulate the results automatically.

The DataSummariser reads in result.xml produced by the TypingAnalyser. The DataSummariser requires each PT/IS pair to only have one resultList. If a pair has more than one resultList, the DataSummariser will give a warning to the user and terminate the process.

The DataSummariser uses the content of <ErrorList> in result.xml, to construct its result, not the content of <resultList>. Therefore, if the user wishes to change the classifications of any error in they must change the content of <ErrorList>.

The DataSummariser produces three CSV files:

- errors.csv - A list of all the errors found. For each typing error it lists participant ID, phrase ID, whether it was corrected or uncorrected, the error type, intended letter and actual letter. This file can be used to analyse individual error types across the whole sample.
- phrases.csv - Lists all PT and IS pairs examined. For each phrases it lists phrase ID, participant ID, PT, IS, TT, number of errors, total time taken, MSD, C, F, IF, INF, TER, CER, WER, CE, PC, UB and WB (for explanation of these terms, please see Section 3.2).
- participants.csv - Lists all the participants examined in the sample, along with their average typing performance statistics. For each participant, it lists - Participant ID, total number of phrases typed, total characters in PT, total characters in IS, total characters in TT, total time taken to type all phrases, mean KSPS (Keystroke per Second), mean WPM (Words Per Minute), mean KSPC (Keystroke per character), mean of MSD, TER, CER, WER, CE, PC, UB, WB and mean accuracy in %.

These .csv files can be imported into programs such as Microsoft™ Excel and SPSS for further analysis.

9.10 Study 1 - Evaluating Accuracy of Categorisation

The aims of the TypingAnalyser were to detect and categorise typing errors and reduce ambiguities. To evaluate the effectiveness of the TypingAnalyser, two studies were devised. The first study evaluated how accurately the TypingAnalyser categorised errors. The second study evaluated how efficiently the disambiguation is carried out by the TypingAnalyser. This section describes the first of these two studies.

9.10.1 Method

9.10.1.1 Design

Raw data collected by the TypingCollector from previous studies were fed into the TypingAnalyser. The results were manually checked for accuracy of the categorisation.

9.10.1.2 Participants

In an effort to evaluate the TypingAnalyser as accurately as possible, all data taken by the TypingCollector available at the time of this study were used (this excluded the pilot study data used in Chapters 5, 6 and 8). In all, the data set contained phrase-copying typing from 412 participants.

Children's typing data consisted of typing from 183 children (Group 3-11, Participant IDs C1-C183) in nine separate one-days studies carried out between 2008 and 2010. The children's ages ranged from 6 to 10 years old (Years 3 to 5). There were 105 boys and 76 girls in the children's group. All child participants were recruited from local schools in Lancashire, representing nine classes in four schools.

Adult typing data was gathered from 229 adults (Group 14 and 15, Participant IDs S1-S229). These were collected from two week-long studies, one in 2008 and the other in 2010. All adult participants were 1st year undergraduate students studying computer science at the author's university. In all, there were 204 male and 24 female participants. Their age ranged from 18 to 44 years old, but the majority (220) were aged between 18 and 30 years old. Further details of the participants can be found in the participant summary chart (Appendix 3).

9.10.1.3 Apparatus

The TypingCollector was used to collect demographic information about the participants, to present the phrases and collect their typing data. Although different keyboards and PCs were used across the studies, all participants were tested on a Windows PC, and they all used full-size QWERTY keyboards with black keys with white writing. All keyboards had a Windows PC UK layout (see Appendix 1 for a diagram of this layout).

9.10.1.4 Procedure

Although the studies varied on what demographic and CE data the participants were asked, the procedure of the phrase-copy typing remained consistent across all studies. Since this evaluation study is only concerned with the data from the phrase-copy typing task, it was deemed that the data is comparable across the various studies.

In all studies the TypingCollector was used to collect personal information about the participants, present the phrases and collect the typing data. Each participant ran an individual copy of TypingCollector (Section 6.5) on his or her PC, and had ten randomly selected phrases from the CPSet (Section 5.2) to copy. The phrases were shown one at a time in font style Verdana at size point 14, with a space for the participant to copy the phrase underneath. The TypingCollector logged what phrases were shown, each key that was pressed, recorded its timestamp and the time between the keystrokes.

9.10.1.5 Analysis

The xml files of raw typing data created by the TypingCollector were analysed by the TypingAnalyser. The classifications of each typing error found were checked manually to see how many were correctly categorised. For typing errors that were not correctly categorised, a note was made as to whether the TypingAnalyser gave a warning or not.

If the participant typed nonsense words, this was removed from the original sample. Nonsense inputs were characterised as a string of random letters, usually with no spaces. These were not considered as misclassifications since differentiating a nonsense input from an erroneous but ‘sensible’ input was beyond the scope of this study. The TypingAnalyser gave warnings of all such nonsense cases.

9.10.2 Results

In total, 3940 phrases, taken from 412 participants (183 children and 229 adults) were analysed. 11 phrases were found to contain nonsense input, so were removed from the sample, leaving 3929 phrases to be categorised.

Table 47 shows how many phrases were correctly categorised by the TypingAnalyser. It shows that the TypingAnalyser achieved 99.6% accuracy in detecting and categorising 3606 typing errors.

Table 47: Accuracy of the TypingAnalyser Detecting and Classifying Typing Errors from 2940 Phrases

	Children	Adults	**Total**
Total number of phrases	1681	2259	**2940**
Phrases containing nonsense input	11	0	**11**
Phrases tested	1670	2259	**3929**
Total number of typing errors detected	2359	1263	**3622**
Number of errors categorised correctly	2350	1256	**3606**
Number of misclassified errors	9	7	**16**

9.10.3 Discussion

9.10.3.1 Accuracy of the TypingAnalyser

This study showed that the TypingAnalyser detected and categorised typing errors at 99.6% accuracy. Where the TypingAnalyser failed to categorise the typing errors correctly, it gave warnings for 12 of the 16 cases. It also gave warnings for all 11 of the nonsense input cases. It is therefore reasonable to say that a user of this analyser will be able to detect any nonsense input and most misclassifications by simply checking those cases that come with warnings, rather than every single case tested, saving considerable time.

9.10.3.2 Limitations

The study was carried out on typing data of 412 participants. Although this is a reasonable sample size, all participants were recruited from Lancashire, UK. The TypingAnalyser is therefore untested in countries with dialectal spelling differences, such as the United States.

The sample also consisted of only young children and computer science students. It did not include teenagers or adult participants relatively new to typing. This means that the sample is likely not to cover the whole spectrum of typing skills available.

9.10.4 Conclusions

The TypingAnalyser performs categorisation of typing errors with very high accuracy. The TypingAnalyser also gave warnings for 12 of the 16 cases of the misclassified errors. This suggests that the TypingAnalyser is a reliable method in classifying typing errors, and can be used instead of manual classifications.

9.11 STUDY 2 – EFFICIENCY OF DISAMBIGUATION

To evaluate how much ambiguity can be solved by the TypingAnalyser, the same sample of typing errors in Study 1 was tested again. In the first instant, the sample was analysed by the TypingAnalyser without any of the disambiguation strategies (Section 9.8). Then a second run with the same data was performed *with* all the new disambiguation strategies.

9.11.1 Method

9.11.1.1 Participants

The same raw typing data collected from 412 participants (183 children and 229 adults) used in Study 1 was used in this study.

9.11.1.2 Procedure

Firstly, the following disambiguation methods were commented out from the TypingAnalyser:

1. assuming only the second letter in a duplicated letter is erroneous - Section 9.4.2
2. deleteTEDuplicates() - Section 9.8.2
3. deleteWordDuplicates() - Section 9.8.3
4. reduceWordTriplets() - Section 9.8.4

Then the raw typing data of 3940 phrases were run through the TypingAnalyser. 11 phrases were found to contain nonsense input, so were removed from the sample, leaving 3929 phrases to be categorised.

The number of cases where there were more than one resultList occurred were counted. The disambiguation methods were put back into the TypingAnalyser. The same data was analysed again, this time with all the disambiguation strategies applied to ambiguous cases. The number of phrases that still contain ambiguities was counted for comparison.

9.11.2 Results

Table 48 below shows how many ambiguities remained unresolved with and without the new disambiguation strategies:

Table 48: Number of Ambiguities Before and After Disambiguation

	Children	Adults	**Total**
Total number of phrases	1681	2259	**2940**
Phrases containing nonsense input	11	0	**11**
Phrases tested	1670	2259	**3929**
Phrases with ambiguities BEFORE disambiguation	175	39	**214**
Phrases with ambiguities AFTER disambiguation	23	2	**25**
Number of ambiguities solved by TypingAnalyser	152	37	**189**

9.11.3 Discussion

The TypingAnalyser was able to resolve 88% of the ambiguities found in the data. This reduces the work of the user from having to examine 214 ambiguous phrases to only 25. This is a considerable amount of time saved on the part of the user. Additionally, the user would be likely to resolve ambiguities in rather an inconsistent manner, whereas the TypingAnalyser followed four strict disambiguation rules to remove duplicate and redundant resultLists.

9.11.3.1 Limitations

Since this study used the same data set as Study 1 (previous) in this chapter, it suffers from the same methodological limitations as Study 1.

9.11.4 Conclusions

The TypingAnalyser was successful in resolving 88% of ambiguities found in the sample data. This means that the TypingAnalyser is able to save considerable time for the researcher carrying out large-scale studies of typing errors. For example, as this study shows, a study of 412 participants typing ten phrases each only required the researcher to manually disambiguate 25 phrases.

9.12 Conclusions

The main contribution of this chapter was the two new tools that automatically detect, categorise and tabulate typing errors found in phrase-copying tasks, with high accuracy and speed. Although automatic categorisation of typing errors existed beforehand, the new TypingAnalyser expanded the categorisation types from 3 to 28 error types, giving far more detail about the context in which the errors occurred. The TypingAnalyser gave very high accuracy in categorising typing errors (Section 9.10.3), and was able to resolve a high number of ambiguities (Section 9.11.3). Additionally, both the TypingAnalyser and DataSummariser carried out tasks that have saved considerable time in carrying out the analysis.

Although the TypingAnalyser was highly accurate in categorising typing errors, it was not 100% accurate. The TypingAnalyser misclassified 16 out of 3622 error types found. Although this equates to only 0.4% of the overall sample, with more time, it is desirable to have this misclassification rate eradicated. The TypingAnalyser also left 12% of ambiguous results unresolved. It is hoped that a further examination of the ambiguous results will shed light on further disambiguation rules that will reduce the number of unresolved cases.

The next chapter will use the TypingAnalyser and DataSummariser, together with the TypingCollector (Chapter 6) to carry out a large-scale study comparing typing errors of children and adults.

10 STUDYING TYPING ERRORS MADE BY CHILDREN AND ADULTS

10.1 INTRODUCTION

With a comparable method of gathering and analysing typing data from children and adults established, it was possible to try and answer the main research question of this thesis - *are there observable differences in the way children and adults make typing mistakes?*

This chapter compares the typing errors made by children with those made by adults. Typing was gathered from 231 children and 229 adults by the TypingCollector (Chapter 6) and typing errors were detected and categorised by the TypingAnalyser (Chapter 9).

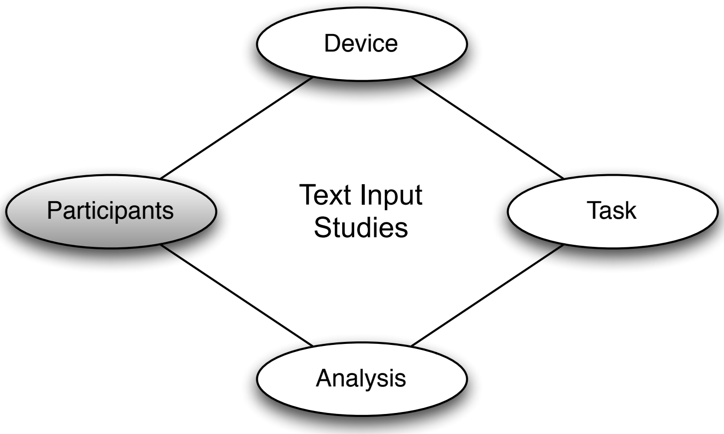


Figure 49: This Chapter Compares the Differences Between the Two Participant Groups

An examination of the error rates defined by Soukoreff and MacKenzie (2003) revealed that there were significant differences in the amount of typing errors the two participant groups made and how they corrected the errors. However, the error rates do not show *how* the two groups differed, or *where* the differences arose.

Their typing errors were categorised using the TypingAnalyser (Chapter 9), which provided more detail about how the two groups differed. Each error type was studied further for differences between how children and adults made typing errors.

The study defined an error to be child-specific if the child participants made more than three times that of the adults, more than one child made the error and no more than one adult made the error. The study found 13 typing error behaviours that fit this category. An error was classified as a child-prone error (errors that children make more of than adults, but adults also made some) the ratio of amount made by children were more than three times greater than those made by adults, but did not fit into the conditions of child-specific errors. The study found nine typing error behaviours that fit this category.

The majority of typing errors specific to children were related to the misunderstanding of the phrase-copying task or the English language, where as child-prone errors were related to actual typing errors. These typing errors show that children make typing errors in a different way to adults. This finding suggests that understanding how adults make typing mistakes cannot be directly applied to children.

10.1.1 Scope

Although there are many ways of analysing typing data, this chapter is focused on the typing errors. The core purpose of the analysis was establishing whether there were differences between the typing errors made by children and those made by adults. Therefore, although brief examination of overall typing speeds and error rates are made at the start, the majority of the comparisons are restricted to the typing errors themselves. Although it is possible to investigate the cause of each error types by studying other features such as the timing of the key presses, this was beyond the scope of this thesis and was were not carried out. Nonetheless, the data collected by the TypingCollector does provide data that will allow for such studies in the future.

Since large numbers of children were required for this study, which meant travelling to several schools, it was more practical for the author to limit the schools to those within Lancashire. It was also easier to collect a large number of 18+ year old participants from university classes. Although this made the adult sample not representative of the general adult population, it was deemed to be a reasonable trade off for large number of participants.

10.1.2 Contributions

The main contributions of this chapter are:

1. *Children's typing error rates significantly differ from those of adults*

   Section 10.3.2 statistically compares the error types of two participant groups and found that they differ significantly. This suggests that there are notable differences between the ways the two groups make typing errors.

2. *A list of the typing errors that are particular to children*

   Section 10.3.4 carries out a detailed investigation of the typing errors made by children and adults. It found that there are certain error types and typing behaviours that are particular to children.

10.1.3 Structure

Section 10.2 discusses the method used in gathering the typing data from participants and the analysis methods used. Section 10.3 discusses the results of the data analysis. Section 10.4 discusses the general trend of certain typing errors being more prone to one participant group than the other. Finally, Section 10.5 concludes with a summary of the findings and discussion for future development of this research.

10.2 METHOD

Typing data of phrase-copying tasks were collected from 231 children and 229 first-year undergraduate students. The phrase-copying task was administered by the TypingCollector (Chapter 6) to ensure the same visual environment for all participants. The same phrase set (CPSet - Chapter 5) was used for both groups.

10.2.1 Participants

For the analysis of typing errors made, most typing data previously collected and some new typing data were merged together to maximise the number of participants analysed. The selected samples were merged on the basis that all the participants carried out the typing task using the TypingCollector. Although the TypingCollector went through several versions during this research, the typing task component remained constant throughout. All the participants typed phrases selected from the same phrase set, displayed in the same font style and size, in the same environment.

Typing data from 231 primary school children aged between 6 and 11 years were used as the children sample. The participants were limited to those that carried out the phrase-copying on the TypingCollector, which were group 3 to 13 (participant ID C1-C231) in the participant summary chart found in Appendix 3. Data from these groups have already been used in previous chapters (Chapter 5 to 9). The age range of the total sample was between 6 and 11 years old but most (207) were between 8 to 10 years old. The child groups represent 13 classes from five schools in Lancashire, UK. All spoke English as their first language.

Typing data from 229 first-year undergraduate computing students were taken on two separate academic years as the adult sample. These two groups are the same groups described in studies in Chapter 9. In both years, the sample was taken on the first semester during their lab time. As with most other computing degree courses, the sample were heavily skewed towards the male population with 204 males and 25 females in total. The age range of the total sample was 18-44 years but most (220) were between 18 and 30 years old. These participants are represented as group 14 and 15, participant ID S1-S229 in the participant summary chart found in Appendix 3.

10.2.2 Apparatus

The TypingCollector (Chapter 6) was used for both the adult and child participants. Although the demographic and computer experience questions varied throughout the research, the phrase-copy typing component remained the same. Although different keyboards and PCs were used across the studies, all participants were tested on a Windows PC, and they all used full QWERTY keyboards with black keys with white writing. All keyboards had a Windows PC UK layout (see Appendix 1 for a diagram of this layout).

10.2.3 Procedure

In all studies the TypingCollector was used to collect demographic and CE information, present the phrases and collect the typing data. Each participant ran an individual copy of TypingCollector on his or her PC, and had 10 randomly selected phrases from the CPSet (Chapter 5) to copy. The phrases were shown one at a time in font style Verdana at size point 14, with a space for the participant to copy the phrase underneath. The TypingCollector logged what phrases were shown, each key pressed, its timestamp and the time between the keystrokes. On average, children took four and half minutes to type the phrases, and adults took on average fifty seconds.

10.2.4 Analysis

Figure 50 below shows the stages taken on analysing the typing data. The xml files of typing data created by the TypingCollector were analysed by the TypingAnalyser. The Initial Result contained error rates discussed in Section 10.3.2, and an xml file of all the errors that were detected and categorised.

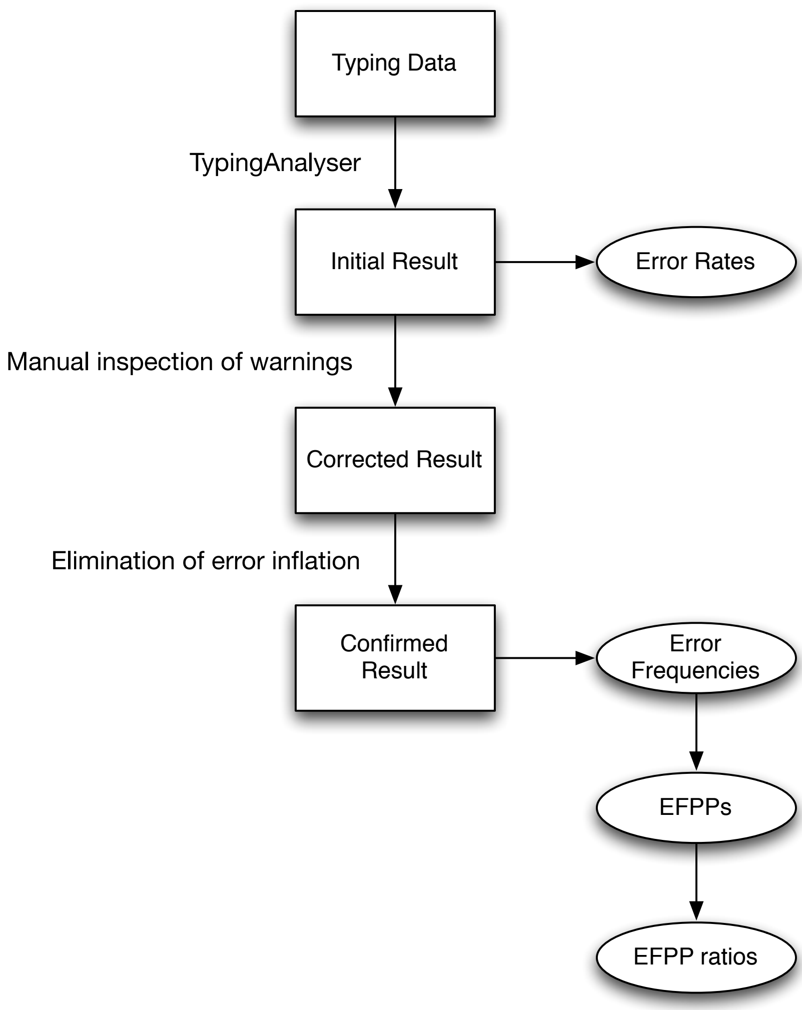


Figure 50: Stages of Analysis Used in This Chapter

This initial result was then manually inspected by the author for warnings given by the TypingAnalyser (see Section 9.9.1). These warnings were manually checked for accuracy of the categorisation and corrected where necessary (0.4% of the results were corrected). This corrected result was summarised by the DataSummariser, which lists data for each typing error, each phrase typed and each participant. The computer experience (CE) score was calculated as defined in Section 7.4.5 and 7.8.5for each child and adult participant.

10.2.4.1 Inflation of Duplicated Spaces

The errors listed in the corrected result were manually inspected for errors that occurred repeatedly in the same phrase. The TypingAnalyser detected 244 Duplicated Spaces (DS, Section 8.5.3.6) from the children. Closer inspection of the errors showed that some of these ‘errors’ had been made purposely. Figure 51 below shows the number of child participants that made certain frequency of DS errors in their typing. 50 participants made 244 DS errors. Although most of them (32 participants) only made one DSs, it shows that eight participants were responsible for 60% (147 occurrences) of DS.

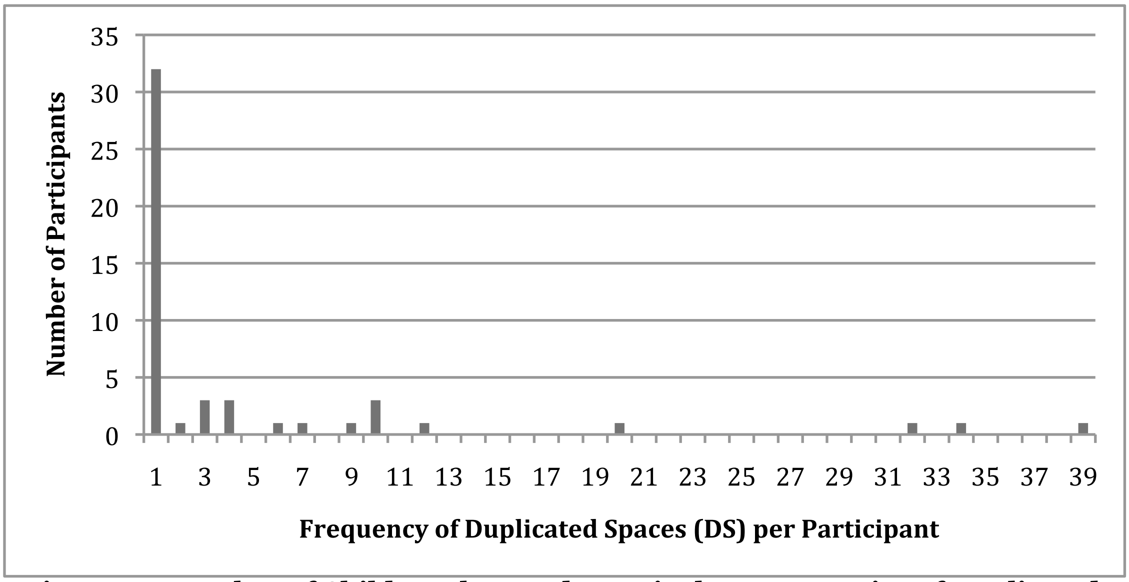


Figure 51: Number of Children that Made Particular Frequencies of Duplicated Spaces (DS)

An examination of the participant who made the most number of DS errors (C168) reveals that the participant consistently duplicated the spaces found in PT (Figure 52 shows the first three phrases typed by C168).

PT_1: `large sack of lamb chops`

TT_1: `large  sack  of  lamb  chops`

PT_2: `sat down on the window seat`

TT_2: `sat  done  on  the  window  seat`

PT_3: `a voice from above`

TT_3: `a  voice  from  above`

Figure 52: First Three Phrases Typed by C168 Who Consistently Duplicated the Spaces Presented in PT

Of the 35 spaces present in PT, 34 were duplicated. This consistency indicates that the child participant *intended* to duplicate all the spaces. All but 2 of the 39 DSs remained uncorrected. Although two DSs were corrected, they were deleted in the process of fixing an error that occurred before the DS. Additionally, these deleted DSs were re-entered by the participant again, and were left uncorrected.

It seems here that the error was not that the participant accidently duplicated the spaces 34 times. It is more accurate to state that the participant made one error in the copying of PT, that they felt *all* spaces should be duplicated. It is therefore inaccurate to count these DSs individually, but to count them as one single error.

However, it is inaccurate to say all DSs made by one participant should be counted as one error. It is possible for a participant to accidently make more than one DS in the course of copying ten phrases. Therefore, it was decided that DS errors should be counted as a single error if at minimum, all but one of the spaces were duplicated. This allows cases where the participant accidently not duplicated an intended duplication, such as seen in phrase C168-8 (`she can drive a train -> she can  drive  a  train`).

Re-examination of the DSs according to this new rule has reduced the total number of DSs from 244 to 100. No adult participants consistently duplicated spaces.

It was assumed in the earlier study of paper-to-screen copying task (Chapter 6) that consistent DS were due to the PT being presented at much larger font size than TT. It was assumed that the children were trying to match the physical size of the gaps between TT words to that seen in PT. However, in this study, PT and TT were of the same font size. Further, the two were displayed on top of each other, meaning that the difference in the size of the 'gaps' was more apparent. It is unclear at this stage as to why the children (although at a reduced frequency) continue to carry out DSs. Consistent duplication of spaces was not observed in adults. In fact, DS only occurred once in the adult group.

10.2.4.2 Inflation of Omitted Spaces

Similar to Duplicated Spaces, some participants made Omitted Space (OS, Section 8.5.1.2) errors on a consistent basis. Two children (C24 and C231) failed to put any spaces between the words typed. It seems likely that they felt copying all the letters in the right order was enough, and the spaces did not matter. In such cases, it was considered to be one error (error in understanding the task) rather than many separate accidental Omitted Spaces.

Another way a consistent Omitted Spaces could be classified was if the participant had to quit the task before he or she had completed all ten phrases. In these cases, the Omitted Spaces were removed altogether.

Dealing with these inflations reduces Omitted Spaces from 219 to 142 errors in children. Although adults did make OS, none made it consistently throughout their typing. It is possible to say that a consistent omission of spaces (or typing without putting any spaces in) is a child specific typing error.

10.2.4.3 Inflation of Capitalisation Errors

A closer examination of Capitalisation Errors (CaE, Section 8.5.2.5) reveals that there were some inflation in the total number of CaE. Some children, either accidently or intentionally, typed with the Caps lock turned on. This resulted in all the typed letters being classified as CaE. An example of this is shown in Figure 53, which shows the first phrase typed by the participant C226. C226 typed all ten phrases in uppercase letters, which were classified as 179 CaEs by the analyser.

PT_4: `I have everything I want`

IS_4: `{CAP_LOCK}I HAVE EVERYTHING {CAP_LOCK}i{CAP_LOCK} WANT`

Figure 53: The First Phrase Copied by C226 - The Participant Forgot to Turn Off the Caps Lock to Type the Remainder of the Phrase

Clearly, this was not 179 separate errors. It is fair to assume that the error was either in not realising that the Caps lock was still on, or thinking that capitalisation of letters were not an error. These causes indicate a lack of awareness of what one was typing, or misunderstanding that copying a phrase meant letter-for-letter but not necessary matching the case.

Whatever the cause, it was decided to count this particular group of CaEs as one error. If all letters after using the Caps lock were uppercase letters, this was counted as one error. A re-examination of the CaEs typed by children revealed that four participants made this mistake. Counting these errors as one error per participant reduced the total number of CaEs from 519 to 164. The evidence that these children did not think this was an error is supported by the fact that of the 348, no attempts were made fix them. Interestingly, there were nine occasions were CaE was deleted by the participants, but this was in an effort to fix an error other than the CaE – implying that these participants did look at what was typed on the screen, but did not feel that the uppercases were an error. Additionally, in all nine cases, the letters were retyped again in uppercase letters. Only one adult participant made a similar error once (S97-9), and this was also counted as one error.

10.2.4.4 Comparing Children and Adult Typing Errors

With the inflated errors reduced, the final confirmed result provided the frequency of each error type. By comparing the amount of errors made by children and adults, the frequencies of each error type for each participant group was standardised to account for the difference in the number of participants and number of phrases completed by the two groups (Error Frequency per Phrase per Participant = EFPP). The EFPP for each error type for each group was calculated as:

$$EFPP = \frac{(F/Ph)}{Pa} \tag{9}$$

Where *F* is the frequency of errors made by the participant group, *Ph* is the total number of phrases that were completed by the participant group and *Pa* is the number of participants in the group. EFPP for children ($EFPP_{children}$) and adults ($EFPP_{adults}$) were calculated for each error type.

In addition to comparing EFPP rates for each error type, individual errors of the same error type were manually examined for any patterns according to the physical locations of the keys involved, and its relation to other errors in the same phrase. By inspecting whether an error type or a typing behaviour is child specific, the ratio of children's EFPP to adults' EFPP ($EFPP_{children}/EFPP_{adults}$) was compared. These are discussed in Section 10.4.

10.3 RESULTS

In total, children typed 2151 phrases, and adults typed 2259 phrases. Children on average (mean) were shown 212 characters, and typed 238 characters. Adults were shown 216 characters on average, and typed 232 characters. The differences in the total number of PT characters shown are due to some children not typing all of the phrases. Children typed a mean number of 9.4 phrases, with the lowest number being 2 phrases (by three participants). The adults typed on average 9.9 phrases, with minimum number of 7 phrases by 1 participant.

10.3.1 Computer Experience Score

The earlier data collection sessions did not have the Computer and Tying Experience Questions (CTEQ) to administer to the participants. The CTEQs were only administered to the 136 adult participants that took part in the second adult data collection (Group 15 on the participant summary, Appendix 3), and 68 child participants (Group 12 to 13, Appendix 3).

Section 7.4.7 showed that the adult participants who had completed the ACTEQ had a positively skewed distribution on how often they used a computer, number of computers and laptops owned, the task score, feeling in control when using the computer and their attitude towards computers. The adult participants showed a normal distribution for the remaining 14 ACTEQ items. Children's answers to their CCTEQ was also analysed in Section 7.8.6 and all CCTEQ items were found to be normally distributed.

10.3.2 Comparing Error Rates

Table 49 shows some of the error rates defined by Soukoreff and MacKenzie (2003) for the children and adults groups (see Section 3.2 for definitions).

Table 49: Children and Adult Error Rates (Defined by Soukoreff and MacKenzie, 2003) with Mann-Whitney U Statistics

	Children (N=231)	Adults (N=228)	Mann-Whitney U	
	Median	Median	U	*p*
Keystroke Per Character (KSPC)	1.080	1.054	59126.0	0.000
Minimum String Distance (MSD)	0.200	0.000	64821.5	0.000
Total Error Rate (TER)	0.040	0.020	67973.5	0.000
Corrected Error Rate (CER)	0.025	0.012	63766.5	0.000
Uncorrected Error Rate (UER)	0.015	0.005	65650.5	0.000
Correction Efficiency (CE)	0.333	0.166	65774.5	0.000
Participant Conscientiousness (PC)	0.333	0.200	60825.0	0.000
Ultilitised Bandwidth (UB)	0.928	0.963	40299.0	0.000
Wasted Bandwidth (WB)	0.072	0.038	65932.0	0.000
Accuracy (%)	95.94	98.03	38176.5	0.000

In all rates shown in Table 49, there were highly significant differences between the children and adult groups. This indicates that as groups, they are significantly different in their typing.

The child group's median MSD is higher, meaning there were larger differences between their PT and TT than those by the adult group. This suggests that children left more typing errors in their TT than the adults did. The TER and UER supports this, both being significantly higher for the children. The children's median TER is twice as high as the adults' indicating the children's typing contained twice as many erroneous characters as the adults' typing did. Children's UER being three times higher than the adults' indicates that children left many more errors in TT than the adults did.

Interestingly, CE and PC tell a different story. Children's CE show that were twice as efficient as the adults in attempting to correct their errors, which contradicts their higher UER rate. Similarly, they scored higher in PC than adults did, indicating that they attempted to correct more errors than the adults did. This contradistinction may imply that although children tried to correct more errors than the adults, they were unable to fix the error correctly.

Analysis of the two groups' typing does indicate that the two groups are statistically significantly different. It suggests that the children attempted to correct more errors than the adults but many more errors remained in children's TT. It also showed that children had a wider variance in their typing performance than the adults did.

However, these error rates do not show *what* those errors were. Do adults just make less of every kind of typing error, or are there some errors that are more likely to be made by the children? These questions can only be answered by studying examples of each error types made by children and adults.

10.3.3 Frequency Comparison of Error Types Between Children and Adults

Figure 54 below shows EFPP of each error types for each participant group with adjustments made for DS, OS and CaE (Section 10.2.4.1) and provides an overview of the difference in frequency between children and adults for each error type. It shows that there are differences between the children and adult group. Most notable are the Substituted Letters, where the children made nearly three times more than the adults did. A general trend can be seen here that there are some error types that children make, but are rarely made by adults ('child prone errors') such as Inserted Spaces, Inserted Functions, Duplicated Spaces, Capitalisation Errors, Omitted Space and Omitted Words). In contrast, NT-Mus were often made by the adults but rarely by children

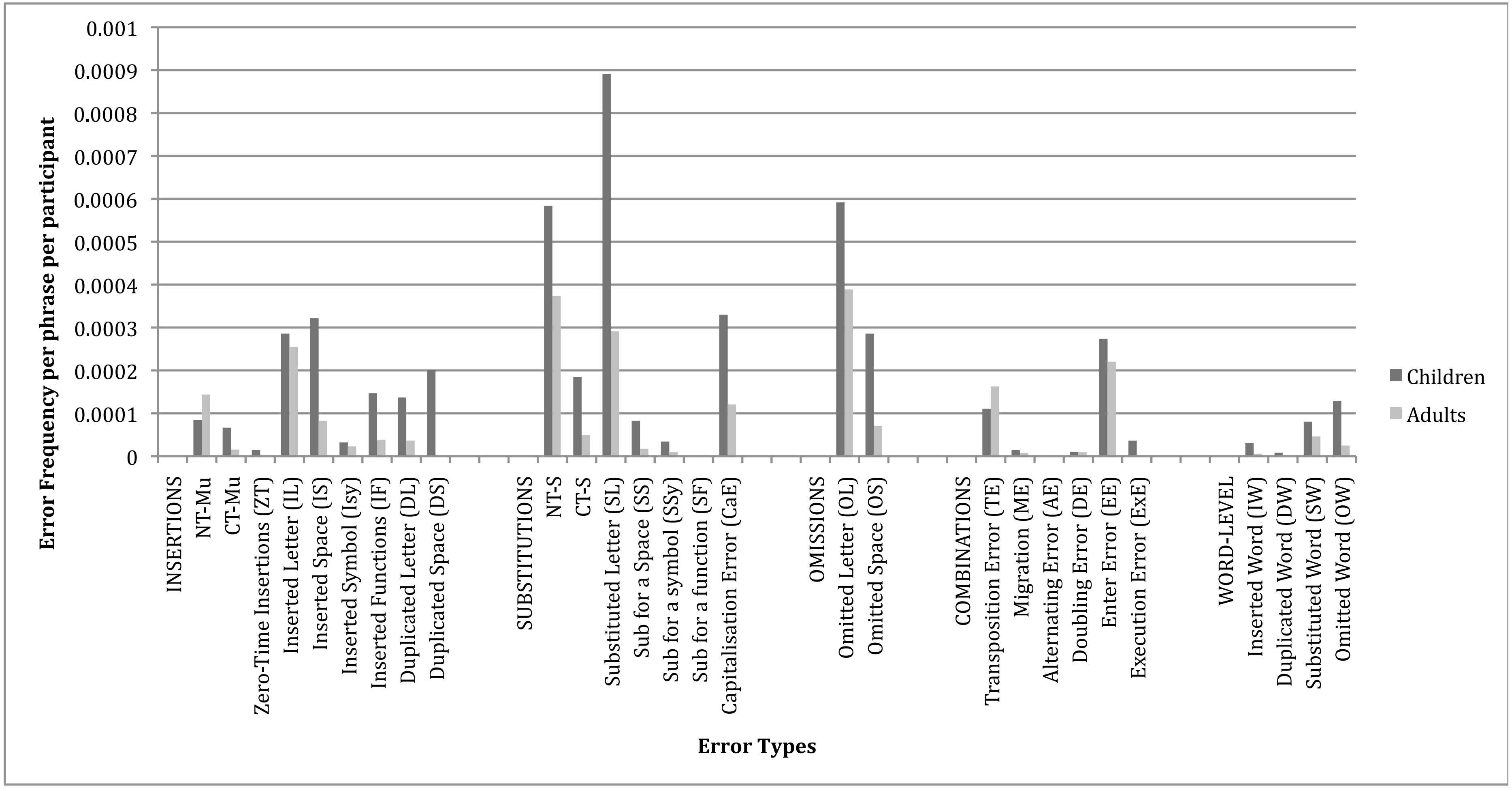


Figure 54: Error Frequency Per Phrase Per Participants by Children and Adults for Each Error Types

10.3.4 Analysis of Each Error Type

Inspection of the error type frequencies does show that there are some differences between children and adults. However, it is when each error type is examined in more detail that differences between the two participant groups really become apparent.

Several error types defined by the ExpECT categorisation method (Chapter 8) have been omitted from this analysis. Firstly, there were no notable differences between $EFPP_{children}$ and $EFPP_{adults}$ of Inserted Letters (Section 8.5.3.1), Omitted Letters (Section 8.5.1.1), Doubling Error (Section 8.5.3.2), Transposition Error (Section 8.5.2.2) and Execution Error (Section 8.8.2). These error types also had no obvious typing error patterns when the errors were examined. Secondly, Alternating Errors (Section 8.5.2.6) and Interchange Error (Section 8.5.2.8) had only one or no occurrences. Finally, none of the phrase-level errors (Section 8.7 were implemented for automatic categorisation by the TypingAnalyser, so were not included in this chapter. There was only one instance of a phrase-level error in the typing sample - a Substituted Phrase where the participant copied the counter at the bottom of the TypingCollector that displays how many phrases they have typed (e.g. '1 out of 10') instead of the Presented Text.

10.3.5 Insertion Errors

Errors involving insertion of letters were the second most frequent errors in both the child and adult groups (26% of all errors made by children and 25% of all errors made by adults). Within letter-level insertion errors, the most common error type made by the children was the Inserted Space, whereas the Inserted Letter was the most common insertion error for adults.

10.3.5.1 NT-Mu, CT-Mu and Zero-Time Insertions

Children made only half the amount of NT-Mu (Section 8.5.3.3) the adults made ($EFPP_{children}$ = 0.000085, $EFPP_{adults}$ = 0.0000144). In contrast, children made considerably more CT-Mu (Section 8.5.3.4) than adults did ($EFPP_{children}$ = .000066, $EFPP_{adults}$ = 0.000015). 45% of the children's CT-Mu involved the space bar, whereas only 25% of the adults' CT-Mu involved the space bar. Although this is interesting, the very low frequency of these errors (14 and 2 respectively where the intended letter was a space) makes it unwise to draw a solid conclusion about the differences between the two groups.

When two keys are pressed simultaneously, the keys can be Next-To (NT, Section 8.5.2.3), Close-To (CT, Section 8.5.2.4) or not adjacent to each other at all. Zero-time Insertions refers to the last group. Children made seven of these errors, where as the adults made none. The details of these Zero-Time Insertions are shown in Table 50.

Table 50: Details of the Seven Zero-Time Insertions Observed

Phrase ID	Intended Letter	Actually typed	PT	IS
C9-4	a	ca	nothing can compare	utyhfihjgij<<<<<<<<<<<<nnnnnn n<<<<<othi<ngt<c**ca**n<<<av<n< <<anCAP_LOCKKJBBJJ<<<<<< CAP_OFF compare
C48-3	g	gu	laughing with the clown	la**gu**<gu<<<uhing with the clown
C64-4	u	uk	I saw a beautiful butterfly	i<CAP_LOCKICAP_OFF saw a beah **uk**huihikgtyklytiooioyijo<<<<<<< <<<<<<<<<<<<<<<<<<<autiful butterfly
C64-4	j	jo	I saw a beautiful butterfly	i<CAP_LOCKICAP_OFF saw a beah ukhuihikgtyklytiooioyi**jo**<<<<<<< <<<<<<<<<<<<<<<<<<<autiful butterfly
C91-5	n	nENTE R	they all go marching down	they all go marching dow**nENTER**
C138-4	i	iSHFT_ RIGHT	I love the rolling hill	**iSHIFT_RIGHT**SHIFT_UP<SHIFT_ IIIISHIFT_UP<<< love thw<e rolling hill
C231-4	o	o.	he kept trotting outside	hh<ekepttrotting**o.**<utside

In 3 of the 7 cases (C9-4 and twice in C64-4), children typed a sequence of random characters. It is probable that the children randomly typed some keys very fast. In three other cases (C48-3, C91-5 and C138-4), the extra key inserted was in fact a key they were supposed to press either immediately before or after the intended key. This is a different kind of typing error to the first three. Here, the timing in which they press the two keys was so close that it registered as Zero-time key press.

The final case, C231-4 is a little more difficult to establish how it occurred. It is likely that the participant intended to type the letter 'o' but also typed '.'. However, the key ‘l’ separates the two keys, so it is not possible for a badly aimed finger to accidently pressed the two keys together. This implies that the participant must have used two fingers (of the same hand or using the other hand) to press the two keys simultaneously. Since children made seven Zero-Time Insertions and adults made none, this error type likely to be a child specific error type.

10.3.5.2 Inserted Spaces

In contrast, children made almost three times as many Inserted Spaces (IS, Section 8.5.3.5) than the adults did ($EFPP_{children}$ =0.00032, $EFPP_{adults}$ = 0.000082). One clarification here is that an Inserted Space is different from Duplicated Spaces (described later), in that an Inserted Space refers to an insertion of a space where no space was intended. Duplicated Spaces refers to an intended space that was doubled. If an intended space was multiplied to more than two, this was classified as an Execution Error (ExE - see Section 8.8.2). Inserted Spaces can be grouped into five types as shown in Table 51 below.

Table 51: Five Categories of Inserted Spaces (IS) Made by Children and Adults

	Children		Adults	
	Frequency	% of total frequency	Frequency	% of total frequency
At the end of a phrase	104	65.0%	32	74.4%
As part of an Inserted Word	34	21.3%	5	11.6%
Splitting a word	9	5.6%	1	2.3%
In middle of a word	9	5.6%	5	11.6%
At the start of the phrase	4	2.5%	0	0%
Total	160		43	

At the end of the phrase - First, and the most frequent of IS, is to insert a space right at the end of the phrase, such as typing '`her father and mother`' as '`her father and mother `'. This was the majority of the IS in both children and adults. Proportionally, adults made 10% more of this type of IS than children.

As part of an Inserted Word - Second type is a space inserted along with a word also inserted into the phrase, such as '`tired of going everywhere`' as '`I am` `tired of going everywhere`'. It is arguable that this perhaps should be counted as part of the Inserted Word error, rather than a separate error. Children made considerably more of this type than the adults did. This is due to the children making more Inserted Words and Duplicated Words (19 in total) than adults (3).

Splitting a word - A third type of IS is when a space is inserted into a long word that the participant may feel it should be two words. For example, the word '`bathtub`' was split into '`bath tub`' twice by the children. It is likely that the participants actually intended on inserting the space, making this an error in spelling (error in intention) rather than in the typing (execution). Since this type of IS was seen nine times in children, but only once by adults, it is likely that the children are more prone to this type of IS than the adults.

In middle of a word - Fourth type of IS was to insert a space in the middle of a word, like in the previous one, but in a way that did not make grammatical sense. An example of this is typing '`is`' as '`i s`'. It is more likely that this insertion was due to a slip of the finger, than intentionally inserting a space like the previous IS type.

At the start of the phrase - The final type was to insert a space right at the start of the phrase, such as typing '`your friends are my friends`' as '` your friends are my friends`'. This was only observed four times in the children and none in the adults: one child (C2) made one error, and another (C161) made three. It is possible that this is a child specific error.

10.3.5.3 Inserted Symbol

For Inserted Symbols (ISy, Section 8.5.3.7) error there were little differences between children and adults ($EFPP_{children}$ = 0.000032, $EFPP_{adults}$ = 0.000023). Here, 'symbols' refer to any keys that are non-alphabetic, but still produce a character on screen (as oppose to function keys that do not produce visible characters). Table 52 below shows details of the 16 ISys made by children. The table also shows whether the intended key and the inserted symbol key were adjacent (NT/CT) to each other or not.

Table 52: Details of Inserted Symbols (ISy) Made by Children and the Locational Relationship between the Intended and Inserted Keys

Phrase ID	Inserted Symbol	Letters before and after	key relation	Corrected
C19-7	3	h and e	e is CT	Corrected
C135-5	4	e and r	both are CT	Corrected
C172-6	4	t and e	e is CT	Corrected
C6-2	-	SPACE and w		Uncorrected
C135-6	;	l	l is NT	Corrected
C89-1	.	at the end of phrase		Uncorrected
C95-2	.	at the end of phrase		Uncorrected
C95-3	.	at the end of phrase		Uncorrected
C95-4	.	at the end of phrase		Uncorrected
C95-5	.	at the end of phrase		Uncorrected
C95-6	.	at the end of phrase		Uncorrected
C2-4	[	SPACE and t		Uncorrected
C229-3	[	w		Uncorrected
C121-1	]	n and SPACE		Uncorrected
C10-3	\	a and v	a is CT	Uncorrected
C55-9	\	SPACE and a	a is CT	Corrected

One child (C95) inserted a full stop at the end of five out of ten phrases they typed. This consistency suggests that the participant *intended* on inserting the full stop, rather than doing so accidently. Another child also inserted a full stop at the end of another phrase. In contrast, no adult inserted a full stop at the end of a phrase. Therefore, putting full stops at the end of phrases is likely to be a child specific behaviour.

Majority of the remaining ISys were due to misaiming of the finger. Six of the remaining eleven ISys involved letters CT to the intended letter. Another ISy involved a symbol NT to the intended letter.

Table 53 shows details of the 12 ISy made by the adult participants. All but 2 (S108-9 and S162-1) of the 12 cases were either NT or CT to each other. Six were NT and four were CT keys, compared to the children's one NT and six CTs. This mirrors the results found in NT-Mu and CT-Mu, where adults made proportionally more NT-Mus and less CT-Mus.

Table 53: Details of the Inserted Symbols (ISy) by Adult Participants and the Locational Relationship Between the Intended and Inserted Keys

Phrase ID	Inserted Symbol	Letters before and after	Key relation	Corrected
S114-9	0	o and f	o is CT	Corrected
S167-3	0	SPACE and o	o is CT	Corrected
S202-2	4	e and r	both are CT	Corrected
S167-10	,	m	m is NT	Corrected
S170-4	,	m and e	m is NT	Corrected
S88-8	,	n and SPACE	SPACE is CT	Uncorrected
S206-10	;	l	l is NT	Corrected
S108-9	[	n and SPACE		Corrected
S162-1	[	e and BACKSPACE		Corrected
S196-4	[	o and p	p is NT	Corrected
S196-7	[	BACKSPACE and p	p is NT	Corrected
S181-9	=	y and BACKSPACE	BACKSPACE is NT	Corrected

It can also be seen that the adult participants did not consistently insert a full stop at the end of the phrase. This suggests that consistently inserting full stops at the end of phrases is a child-specific error. A further contrast between the two groups comes from Table 52 and Table 53. Children corrected only 38% of their CT-Mus, whereas adults corrected 92% of their CT-Mus.

10.3.5.4 Inserted Functions

Children made almost four times as many Inserted Functions (IF, Section 8.5.3.8) as adults did ($EFPP_{children}$ = 0.00015, $EFPP_{adults}$ = 0.000038). In terms of frequency, children made 73 and adults made 20 Inserted Functions. Table 54 shows the frequency at which each function key was pressed by both participant groups.

Table 54: Frequency of Inserted Function (IF) Errors by Both Participant Groups

Function Key	Frequency - Children	Frequency - Adults
CAPS_LOCK	36	2
SHIFT_RIGHT	10	7
CTRL_RIGHT	9	5
INSERT	5	1
SHIFT_LEFT	4	2
ALT_RIGHT	4	0
CTRL_LEFT	2	1
F12	1	0
HOME	1	1

PAGE_DOWN	1	0
DELETE	0	1
Total	73	20

The most frequently inserted function for the children was the Caps lock. This was mostly inserted when they were attempting to insert an uppercase letter when a lowercase letter was required. Table 55 below shows the number of participants that used the Caps lock and the Shift keys to capitalise single letters (both intended and unintended) in both participant groups.

Table 55: Number of Participants that Used Caps Lock and Shift Keys to Capitalise Single Letters

Keys Used	Number of Child Participants	Number of Adult Participants
Caps lock	49	15
Shift	16	72
Both Caps lock and Shift	4	0

For children, the Caps lock was used more frequently than the Shift to capitalise a single letter. 49 children did this, and some made more than one capitalised letter by the use of the Caps lock. In comparison, the adults used the Shift key more often. Therefore, if a Caps lock is used, the participant is more likely to be a child. In particular, the use of Caps lock by continuously holding it down whilst searching for the intended letter (as you would do when using Shift to capitalise) is particular to children's typing. Additionally, mixing the use of both Caps lock and Shift was only seen in children.

10.3.5.5 Duplicated Letters and Duplicated Spaces

The children made over three times as many Duplicated Letters (DL, Section) as the adults did ($EFPP_{children}$ = 0.00014, $EFPP_{adults}$ = 0.000038). The letters 'l' and 'o' were most frequently duplicated by the children at 12 times each, closely followed by the letter 'e' at 11 times. This was not the same for the adults, where the letters most frequently duplicated was 'r' (4 times) and 'l' (3 times). However, the low frequency of each letter means that it is not possible to draw a positive conclusion as to duplication of any particular letter to be a child specific error type.

As discussed in Section 10.2.4.1, children made many more Duplicated Spaces (DS, Section 8.5.3.6) than the adults did. Children made 244 DSs, compared to just one by the adult. Even with consistent duplication of letters removed, children still made nearly ten times as many DSs as the adults did ($EFPP_{children}$ = 0.00020, $EFPP_{adults}$ = 0.0000019). It is clear from this sample that DSs are a very child specific typing error.

10.3.6 Substitution Errors

Errors involving the substitution of letters were the most frequently made errors for both the child and adult groups (42% of all errors made by children and 36% of all errors made by adults). The most common substitution error for the children was the Substituted Letter, whereas for the adults it was NT-S.

10.3.6.1 NT-S, CT-S and Substituted Letters

For Next-To Substitution (NT-S, Section 8.5.2.3) errors $EFPP_{children}$ = 0.00058, $EFPP_{adults}$ = 0.00037. Almost all the letters of the alphabet were substituted for another letter in both groups. In a sharp contrast to NT-S, adults made far less Close-To Substitutions (CT-S, Section 8.5.2.4) than children ($EFPP_{children}$ = 0.00019, $EFPP_{adults}$ = 0.00005).

Although there were 15 cases (out of 92) where the children substituted a space for a CT key, adults made no CT-S (out of 26 cases) of this kind. Additionally, 24 CT-S by children involved letters substituted to a space, but adults made only four CT-S of this type. This may mean that a CT-S where either the intended or the actually typed letter is a space is likely to be a child-prone error type.

Children made three times as many Substituted Letter (SL, Section 8.5.2.1) as the adults did ($EFPP_{children}$ = 0.00089, $EFPP_{adults}$ = 0.00029). SL was the most frequent error type made by children (18% of all errors made by children), but not by adults (most frequent errors made by adults was Omitted Letter).

10.3.6.2 Substituted for a Space

A similar trend is found in Substituted for a Space (SS), where the children made four times as many SS as the adults did ($EFPP_{children}$ = 0.000082, $EFPP_{adults}$ = 0.000017). Children substituted other letters into a space 41 times where as the adults only did this nine times. Although SS cannot be said to be a child-specific error type, it is likely to be a child-prone error type.

10.3.6.3 Substituted for a Symbol

Children made three times as many Substituted for a Symbol (SSy) as the adults did ($EFPP_{children}$ = 0.000034, $EFPP_{adults}$ = 0.000009). Table 56 below shows details of the SSys made by children.

Table 56: Details of the Substituted Symbol (SSy) Errors Made by Children

Phrase ID	Intended	Actually Typed	PT	IS	Key Relations
C5-2	a	-	needed a longer piece of string	neededa -< <<< a longer piece of string	
C6-8	h	¬	how I wonder what you are	how SHIFT_RIGHTISHIFT_UP wonderw¬ << what you are	
C9-10	i	¬	in the month of June	¬lkioljko<<<<<<<<<j<in the month of june	
C28-8	y	=	they sailed away for a year	thee=<<y sailed away for a year	NT to enter
C54-1	t	9	it began to rain	i9<t began to rain	
C64-1	SPACE	8	a going home present	98g8t<<<<<<<<a going home present	
C64-1	a	9	a going home present	98g8t<<<<<<<<a going home present	
C64-1	o	8	a going home present	98g8t<<<<<<<<a going home present	
C67-2	l	=	I like green frogs	ENTERENTER<CAP_LOCK ICAP_OFF<<<<=<<<<<<< <<<<<<<<<<<<<<<<<<<< <<<<like geen frogs	NT to enter
C67-5	l	1	laughing with the clown	1<i<laug <hing m<with tn<he clown	
C71-8	SPACE	,	he was nothing but a pup	he ,<was <nothing but a pup	CT to SPACE
C82-1	SPACE	=	the boat rocked gently	the ba===<<<<oat rocked gently	NT to enter
C82-1	r	=	the boat rocked gently	the ba===<<<<oat rocked gently	NT to enter
C82-1	t	=	the boat rocked gently	the ba===<<<<oat rocked gently	NT to enter
C173-8	f	]	they sailed away for a year	theysailer<d away]<<<<<<<< <<<< sailed away s<foe<r a year	
C175-8	e	=	bake me a cake	bake me a caj=<<ke	NT to enter
C229-2	d	4	the mayor held up a medal	the mayor hei<ld upa< aCAP_LOCK M<ME4<<<CAP_OFFmef<dal	CT to e

Most interesting of the SSy are those that involved the letter being substituted for the symbol '='. In all cases, the participants made a typing error, and was intending on pressing the Backspace to correct it. However, they all pressed the '=' instead, which is NT to the Backspace.

It is unclear whether C64-1 typed some random letters (ignoring the PT), or made many typing errors whilst trying to type PT. Some of those numbers are CT-keys to some of the first few letters of PT. Table 57 shows details of the SSy made by the adult group.

Table 57: Details of SSys Made by Adults

Phrase ID	Intended Letter	Actual Letter	PT	IS	Key Relations
S30-2	I	\	I know what to do	\SHIFT_RIGHTISHIFT_UP<< SHIFT_RIGHTISHIFT_UP know what to doENTER	
S38-10	t	=	the fairy godmother was right	teh <<<he fairy godmother was righ<g=ht<<<<<ght	NT to enter
S66-1	e	,	the cat pleased me	the cat pleased m,<e	NT to m
S178-4	o	,	on the compost heap	on the com,<m,<<post heap	NT to m
S178-4	p	,	on the compost heap	on the com,<m,<<post heap	NT to m

Of the five cases by adults, in four cases (S38-10, S66-1 and twice in S178-4) the intended and actually typed keys were NT or CT to each other. As with the children, in the case involving the symbol '=', it was pressed instead of the Backspace.

In all three cases where a letter was substituted for the symbol ',', they were all typed after the letter 'm', which is CT to the comma. It is likely that this was some sort of a CT error, although the two keys were not pressed at the same time, so cannot be classified as CT-Mu. Only adults made this error.

10.3.6.4 Capitalisation Error

Children made three times as much Capitalisation Error (CaE, Section 8.5.2.5) $EFPP_{children}$ = .00033, $EFPP_{adults}$ = .00012). In both groups, the most frequent CaE was typing a lowercase 'i' for an uppercase 'I' (73% of children's CaE and

92% of adults' CaE). However, it is perhaps unwise to draw conclusions about how each participant group behave regarding typing ‘i’ for ‘I’, since not all participants were shown phrases that contained the word ‘I’. In adults, there were only 3 (out of 63) CaE involved letters other than ‘I’. In contrast, children made 45 such CaEs (out of 164). This suggests that children are more likely to make a CaE that involves the letters other than ‘I’.

Two children (C52 and C95) capitalised the first letter of 4 out of 10 phrases each, none of which were corrected. C52 capitalised on the first four phrases, whereas C95 capitalised the first three and the fifth, but not the fourth. Since they were never told to capitalise or not to capitalise (they were told to do whatever they felt), the two children stopping capitalising is not due to any instruction that was given to them.

Of the 45 cases of children’s CaEs that did not involve the letter ‘I’, 21 were capitalisation of the initial letter. It is likely that capitalisation of the first letter of a phrase is child specific (this was not seen in the adult group). However, none capitalised the initial of *all* the phrases. No adults participant made more than one CaE error, it is possible that consistent capitalisation of the first letter of the phrase is a child specific error.

Of the remaining 24 CaEs by children, 8 were uppercase letters (such as 'June') reduced to a lowercase letter, and 16 were lowercase letters that were capitalised. In 9 of the 16 cases where lowercase letters were capitalised, the Caps lock key was pressed when the participant pressed the 'a' key (Next-To). It may be more suitable for these errors to be classified as NT-S. In three further cases, the participant forgot to take the Caps lock off after an intended uppercase letter.

10.3.7 Omission Errors

Children made 219 Omitted Spaces (OS, Section 8.5.1.2), whereas adults made only 37. After removing the consistent OS (Section 10.2.4.2), this number is still at 142 for children ($EFPP_{children}$ = 0.00029, $EFPP_{adults}$ = 0.000071). Since $EFPP_{children}$ is four times that of $EFPP_{adults}$, it is probable that OS is a child prone error-type.

10.3.8 Other Letter Level Errors

10.3.8.1 Migration Error

Children made twice as many Migration Errors (ME, Section 8.5.2.9) as adults did ($EFPP_{children}$ was 0.000014, $EFPP_{adults}$ was 0.0000076). However, these numbers are based on very low frequencies of occurrence (seven for children, four for adults) and so care must be taken in drawing conclusions from this.

10.3.8.2 Enter Error

Children made similar amounts of Enter Error (EE, Section 8.8.1) to what the adults made ($EFPP_{children}$ = 0.00027, $EFPP_{adults}$ = 0.00022). Children made 136 EEs, majority (128) of which were at the end of the phrase. It is likely that they pressed the Enter key expecting the software to move onto the next phrase. However, these EE were not limited to the first phrase each participant typed. Table 58 below shows details of the eight EEs that did not occur at the end of the phrase, but either at the start or in the middle of a phrase.

Table 58: Details of Enter Errors (EE) Made by Children that Involved the Enter Key

Phrase ID	Location of Enter	PT	IS
C67-1	Middle	I heard you whispering	iENTERENTER<i heard<b<d yoCAP_LOCKCAP_OFF <<<<<<<<< head you whispering
C67-2	Start	I like green frogs	ENTERENTER<CAP_LOCKICAP_OFF <<<<=<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<< <<like geen frogs
C114-2	Middle	that noisy thing	that noisytingENTER<<<< thihngENTERENTER<
C156-1	Start	you look tired	ENTERyou look tired
C168-4	Middle	he rolled around in the mud	he rolled around imENTER<n the mud
C189-7	Middle	they were playing princesses	they were playing i<princessENTERes

In the cases of C67-1, C114-2 and C168-4, the participant entered the wrong character. To correct these errors, they reached for the Backspace key, but pressed the Enter key instead, which was just below the intended key. This is in fact a CT-S error, and perhaps should be classified as so, rather than as an Enter Error.

The adults made similar amount of EEs but they were all at the end of the phrase. It is possible that pressing the Enter key at the start or the middle of the phrase is a child specific error. However, care should be taken to draw this conclusion since half of these were in effect CT-S, which the adults have also made in numbers (see Section 8.5.2.4).

10.3.9 Word-Level Errors

Children made more of all four types of word-level errors found than the adults did. In total, children made three times as many word-level errors as the adults did ($EFPP_{children}$ = 0.00025, $EFPP_{adults}$ = 0.000077).

10.3.9.1 Inserted Word

Children made five times more Inserted Word (IW, Section 8.6.3) than the adults did ($EFPP_{children}$ = 0.00030, $EFPP_{adults}$ = 0.000057). Children made Inserted Word (IW) errors 15 times. However, 11 of these IWs were made by one participant (C73) who altered all the phrases shown (e.g. ‘`we could do a better job`’ became ‘`we could do a better job at marcs and spensers`’). Therefore, it is reasonable to consider this person to be an exception, and consider the total of children’s IWs to be four. Of these, two of these words could be found in previous phrases shown to the participants, and one appeared earlier in the phrase itself (see Table 59 below).

Table 59: The Four Inserted Words (IW) Made by Children (Source Keys: T = This Phrase, A = Another Place (Previously Shown to the Participants), U = Source Unknown)

Phrase ID	Inserted Word	PT	IS	Source
C114-8	of	through all kinds of weather	thy<rough all of kinds of weather	T
C156-2	you	would like to sail the ocean	would you likr<e to sail the oacen<<<<cean	A
C191-1	am	you are very clever	yes i am very clever<<<<<<<<<<<<<<<<<<<<you are very clever	U
C215-3	the	through all kinds of weather	through all the kinds of weather	A

In the first case (C114-8), the child had inserted the word '`of`' earlier than it was intended, and then typed it again at the correct place in the phrase. In the second case (C156-2), the child altered the phrase into a question. It is possible that the child misread the phrase as a question. The inserted word '`you`' does appear as the first word in the phrase shown to the participant one before this phrase ('`you look tired`'). It is notable that the typed phrase is still grammatically correct.

In the third case (C191-1), the child typed a reply to the shown phrase rather than copy it. This is very interesting, as it indicates that the child forgot the task (copying of the shown phrase) and responded to the phrase instead. This is the only case where this occurred. The child, however did notice their mistake, and deleted what they typed to copy the phrase properly.

Finally (C215-3), the participant inserts the word 'the' in the phrase. The word '`the`' appears in the previous phrase shown to the participant ('`standing in the hall`').

Of the words inserted by adults, the sources were unknown for all three words. Table 60 below shows the details of these IWs. In both children and adults, the words inserted were all very short words, mostly made of two letters.

Table 60: The Four Inserted Words (IW) Made by Adults (U = Source Unknown)

Phrase ID	Inserted Word	PT	IS	Source
S62-6	is	this cold and frosty morning	this is <<<cold and frosty morning	U
S100-2	in	it stings the toe	it stings in the toe	U
S150-8	of	pulled out the plug	pulled out of the plug	U

10.3.9.2 Duplicated Word

Four Duplicated Words (DW) were made by children, and none were made by adults ($EFPP_{children}$ = 0.000008, $EFPP_{adults}$ = 0). This implies that DW is likely to be a child specific error. Although it seems obvious to assume that the children made DWs because they had difficulties tracking their place in PT, a look at the age of these participants reveal that they were not the youngest participants in this study. All were aged between 9 and 10 years and were in Year 5. They came from three different schools.

10.3.9.3 Substituted Word

Children made 40 Substituted Word (SW, Section 8.6.2) errors, where as the adults made 24 SW errors ($EFPP_{children}$ = 0.000080, $EFPP_{adults}$ = 0.000046). Of the 40 SWs made by children, in two cases, the intended word was substituted by a word that had appeared in previous phrases. In three other cases, the typed words were found in the current phrase itself. In the remaining 17 cases, the source of the typed words was unknown.

It is not possible to figure out what word the participant intended to type if they corrected their SW immediately. Therefore the study of the phrase these participants typed are restricted to SWs that remained uncorrected. These uncorrected SWs are listed below in Table 61.

Table 61: Uncorrected Substituted Words (SW) Made by Children

Phrase ID	Intended Word	Word Typed	PT	IS
C53-3	a	the	we spent the night in a tree	we spent the night in the tree
C63-6	fetched	fetch	fetched a saucer of milk	fetch a su<aucer of milk
C73-10	poor	pore	give the poor dog a bone	give the pore dog a bone sead mum
C75-8	the	a	hid under the bridge	hid ung<der a bridge ENTERENTER
C79-4	five	5	five little speckled frogs	5 little speckled frogs
C83-6	to	at	looked up to the stars above	look <ed up at the <stars above
C103-9	things	toys	a bag full of things	a bag full of toys
C117-3	till	to	till the moon grew dim	to the moon grew dim
C132-6	our	are	she could be our friend	she coul be are frinds
C151-1	the	a	bumped on the log	bumped on a la<og
C157-4	bake	back	bake me a cake	back ma<e a cake
C167-6	on	at	sat down on the window seat	sat down at the wim<ndow seat
C168-2	down	done	sat down on the window seat	sat done on the window seat
C168-6	down	done	down by the pond	done by the pond
C168-10	down	done	down on my knees	done on n<my knees
C184-4	a	the	give a dog a bone	give the dog a bone
C197-2	looking	look	looking for a star	look for a star
C205-2	these	the	under these nettles	under the nettels

Almost all the phrases still made grammatical sense after the substitution. In some cases, the phrase kept its meaning, whilst others had slightly altered meaning. The most common word substitutions occurred between '`a`' and '`the`' (four times). There were two cases (C63-6 and C197-2) where the intended word were shortened, such as '`fetched`' to '`fetch`' and '`looking`' to '`look`'.

C168 was shown the word '`down`' in three separate phrases. In all three cases, C168 typed it as '`done`'. The consistency of this SW suggests that this child spells '`down`' as '`done`' normally. However, it is not possible to determine the accuracy of this assumption without further testing.

Like the children, majority of (18 out of 24) adults' SWs involved typed words that did not appear in PT. Only four appeared in a previous phrase and further two appeared in the current phrase. Table 62 below lists all the uncorrected SWs made by the adults.

Table 62: Uncorrected Substituted Word (SW) Made by Adult Participants

Phrase ID	Intended Word	Typed Word	PT	IS
S3-9	it	this	make it go away	make this d<go awaty<
S11-8	bringing	bring	bringing the fishermen home	bring the gfis<<<<fishermen home
S11-9	there	the	there were lots of fairies	the wr<ere lots ofgf< << fairies
S30-6	young	you	a strong young man	a strong you man
S46-10	bake	back	bake me a cake	back me a cake
S88-7	foxes	fox	the young foxes ran off	the young fox ran off
S95-10	raincoat	coat	taking off his raincoat	taking off his coat
S96-7	a	the	here comes a fish	here comes the fish
S97-1	the	a	hid under the bridge	hid under a bridge
S107-1	the	a	she got the vacuum cleaner	she got a vacuum cleanerENTER
S132-9	hand	head	my hand on my head	my head on my head
S141-9	June	july	in the month of Jun	in the month of july
S179-7	took	take	they took some honey	they take some honey
S217-5	climbed	climbing	climbed up on top	climbing up on top
S220-10	to	a	would you like to go	would you like a go
S226-2	bone	home	give the poor dog a bone	give the poor dog a home

The SWs observed in the adult sample shows similar behaviour to the children. The substitution between '`a`' and '`the`' is still common (3 times). Shortening of the intended word is observed more frequently than in the children sample (5 times). Most phrases still made sense after the insertion. Some of the phrases retained its original meanings, whereas others had slightly altered meanings. These tables show that there were no notable difference in the way children and adults substituted words.

10.3.9.4 Omitted Word

The most common word-level error type made by the children was Omitted Words (OW, Section 8.6.1). Children made five times more OW errors than the adults did ($EFPP_{children}$ = 0.00013, $EFPP_{adults}$ = 0.000025). The most common word omitted by the children was the word 'the' (10 times). Table 63 shows words that were omitted more than once in the child sample. Adults did not have such a clear pattern of omitted words, with only two words omitted more than once ('`the`' and '`up`').

Table 63: Frequency of Words Omitted (OW) by Children and Adults More than Once

Word	Frequency of Omission by Children	Frequency of Omission by Adults
the	10	2
is	3	0
it	3	0
on	3	0
a	2	0
by	2	0
last	2	0
light	2	0
of	2	0
up	1	2

10.4 DISCUSSION

A close examination of the typing errors, categorised into error types has revealed that some errors are highly specific to children. The following sections identify typing error behaviours that are specific to children (those that were almost solely made by children) and child-prone (those that were mostly made by children but some adults also made).

10.4.1 Child-Specific Typing Error Behaviours

A decision was made by the author to define a typing error behaviour as child-specific where the ratio of children's EFPP to adults' EFPP ($EFPP_{children}/EFPP_{adults}$) had to be greater than 3.0. Additionally, more than two child participants had to make the error, and no more than one adult participant made the error.

Here are the 11 typing behaviours almost exclusive to the children with the EFPP ratio indicated in brackets where adults made more than one of the specified error:

- Consistent duplications of spaces throughout the phrase (144 times by children, none made by adults)
- Consistent omission of spaces throughout the phrase (77 times by children, none made by adults)
- Consistent capitalisation of letters (355 letters by children, none made by adults)
- Consistent capitalisation of the first letter of the phrase (21 times by children, none made by adults)
- Enter Error at the beginning or middle of phrase (6 times by children, none made by adults)
- Zero-Time Insertions (7 times by children, none made by adults)
- Mixing use of Caps lock and Shift keys to capitalise letters (4 times by children, none made by adults)
- Duplicated Words (4 times by children, none made by adults)
- Execution Errors (18.90)
- Consistent insertion of full stops at the end of the phrase (6.25)
- Inserted Spaces (3.90)

Of these, many were due to misunderstanding the task of copying the phrases. Consistently capitalising the initial letter suggests that their training in the classroom of always capitalising the initial letter overrode the need for copying the phrase just as it was shown. A similar explanation is likely for consistently inserting a full stop at the end of each phrase. Although these are not typing errors, they are very specific to children's typing.

Omission of all spaces in copying PT indicates that the participants misunderstood the copying task as only being relevant to copying the letters and not the spaces. Typing all their phrases in capital letters show that the participants felt that as long as the letters matched with PT, it did not matter that their cases did not.

Holding down the Caps lock key to capitalise a letter indicates that the participant misunderstood how Caps lock worked. They understood that pressing the Caps lock would produce an uppercase letter, but not that they could let go of the key. Instead, the participant used the key in a similar way to using the Shift key, which has to be held down to produce an uppercase letter. Although not an error, mixing the use of Caps lock and the Shift key in capitalising letters by a single participant appears to be a child-specific behaviour. Adults in the study appear to have an established preference of which to use, and none mixed the two keys.

In earlier studies of paper-to-screen copying tasks that used different font sizes for PT and TT, it was assumed that children consistently inserted more than one space to match the size of the much larger spaces in PT. However, it appears that there is another reason why children do this, since the behaviour was still found in this study where PT and TT were of the same font size and displayed on top of each other.

Splitting a long word (such as '`godmother`') into two words ('`god mother`') by the children seems to indicate a misunderstanding in the grammar. Another child specific use of the space was to inserting it at the start of the phrase. It was unclear from the examples of this type of Inserted Spaces as to the cause of this.

10.4.2 Child-Prone Typing Error Behaviours

Besides the child specific typing error behaviours, there were errors that children made many more than adults. These error types are therefore child-prone errors, more likely to be made by a child but does not rule out the adult making the same mistakes. A decision was made by the author to define an error as child-prone where the ratio of children's EFPP to adults' EFPP ($EFPP_{children}/EFPP_{adults}$) was greater than 3.0. These errors are listed below, with the EFPP ratio indicated in brackets:

- Substituted Symbols that is not NT or CT (12.49)
- Capitalisation Error that involves a letter other than 'I' (9.37)
- Omitting the word 'the' (5.20)

- Substituted for a Space (4.78)
- CT-Mu (4.33)
- Omitted Spaces (4.03)
- CT-S (3.76) - In particular SPACE to a CT letter and CT letter to SPACE
- Word level errors (3.23)
- Substituted Letter (3.06)
- Using Caps lock to capitalise single letters (3.04)

In a contrast to the child specific errors, the child-prone errors appear to be more due to errors in the act of typing. Children in this study appear to be more prone to errors involving keys that are CT to each other (CT-Mu and CT-S) than the adults were. Substituted for a Space, Substituted Symbols (that is not NT or CT) and Omitted Spaces are also likely to be cause by an error somewhere in the typing process.

Capitalisation Error of letters other than '`I`' were mainly due to three factors. First was the consistent capitalisation of the initial letters, which was due to error in understanding the task. Second was the failure of initialising uppercase letters such as '`June`'. Third was mostly due to accidently pressing the Caps lock when typing the NT letter '`a`'.

71% of the children that capitalised a letter used Caps lock, where as only 17% of adults did. The use of Caps lock to capitalise a letter is not strictly an error, but is certainly a child-prone typing behaviour.

Word-level errors were mostly due to failure to follow PT whilst carrying out the phrase-copying task. In particular, errors focused around words that only changed the meaning of the phrase only slightly, such as the word '`the`'.

10.4.3 Changes to Error Type Classifications

This chapter examined typing errors that were classified into error types by the TypingAnalyser. Each error type was studied in detail by grouping the errors even further according to observable features such as the letter involved or the physical location of letters on the keyboard. It was discovered that some errors were more suitable to be classified as another error type.

Section 10.3.6.4 showed that a Capitalisation Error that was due to the participant pressing the Caps lock key when they intended on pressing the letter 'a' should be classified as an NT-S error. Section 10.3.8.2 discussed how an Enter Error that is caused by the user intending on pressing the Backspace key to correct an error but pressed the Enter key instead, which should be classified as a CT-S. Section 10.3.5.2 highlighted that Inserted Space that is added due to extra words being inserted should be counted as part of the Inserted Word error, and not as an individual error on its own.

Although their original classifications are not incorrect, it was also not accurate in describing the error. The TypingAnalyser requires further development to takes these misclassifications into account.

10.5 Conclusions

This study compared the typing made by children and adults by categorising their typing errors. Although the error rates indicated that there were statistical differences in the rate at which they made typing mistakes and corrected them, it was unclear as to how these differences occurred.

Without automatic categorisation, error rates were the choice of analysis since they were considerably easier to compute. However, using the TypingAnalyser introduced in this thesis has made analysis of typing errors considerably more accessible. The error types offer far more detail than the error types and make it easier for researchers to find trends such as consistent Duplicated Spaces (Section 10.2.4).

The detailed analysis of error types revealed a set of typing error behaviours that were specific to children. The majority of these error behaviours were due to misunderstanding of the phrase-copying task itself. Only three error types were due to errors in the task themselves. In contrast, the majority of the child-prone errors were due to an error somewhere in the typing process, and in following PT during the typing.

10.5.1 Limitations

At this point, it is important to realise that although the adult participant group did not make these error types, it does not mean that adults would never make one. Similarly, because a particular child-specific or child-prone typing error is found in a person's typing, does not automatically imply that the participant is a child. An existence of one child-specific typing error can only suggest that the typist was more likely to be a child than an adult.

Additionally, some children made no child-specific or child-prone errors. This means that studying of one person's typing to determine whether they are a child or an adult is not yet possible.

Some typing errors were classified as one typing error but a closer inspection indicated them to be of another error type. Although the original classifications were not incorrect, adjustments in the TypingAnalyser should be made to handle these cases.

There were many more child specific and prone errors and behaviours than were found for the adults. It is entirely possible that many more adult-prone typing behaviours will be found if a systematic and thorough search of typing errors and behaviours was carried out. However, since the purpose of this thesis was to establish observable differences between how children and adults make, a complete list of child and adult specific typing errors was beyond the scope of this thesis.

This study was based on the typing of 231 children and 229 adults, all selected from the Lancashire area of England. To generalise these findings and to eliminate any error types that may be geographically specific, a larger study involving many more participants, selected from all over the UK is required. Furthermore, the 'adult' sample were all collected from first year undergraduate computing students, who are likely to be very proficient at typing. The adult sample should be expanded to a much wider range of age and typing experience. If possible, samples should also be collected from other English speaking countries, such as America, Canada and Australia, to establish whether these differences are applicable to all English-speaking typists.

There are two limitations to the CE scores. Firstly, the CTEQs were only administered to some of the participants (68 out of 231 children and 137 out of 229 adults). Therefore, the mean CE scores and its distributions do not represent the whole sample. Additionally, since all participants in the adult sample were computing students, it is highly likely that their mean CE score was higher than it would have been if a more representative adult sample (with mixed background in computer use) were selected for the study.

10.5.2 Conclusions

The study established that there is a difference between how children and adults make typing mistakes, in the form of sets of typing errors that each participant group was more prone to. However, it is not yet clear as to why these differences occur. Additional psychological tests to measure the participant's cognitive and motor control abilities may be required to establish what aspects of cognitive and motor-skills that dictates each typing error, and whether or not specific ones are prone to be made by children.

11 CONCLUSIONS

This chapter reviews the research questions the author set out to answer, and to what extent this was achieved. It highlights the major contributions this thesis has made, and puts these in context with works by other researchers. Discussion on the future development of the research is made, before the conclusion is drawn on this thesis.

This thesis set out to answer the question '*are there any notable differences between typing errors made by children and adults?'* Other objectives of the thesis were to establish a comparable typing data collection method between children and adults, establish a typing error categorisation method that encompassed the whole range of typing errors made by both children and adults without making assumptions to their cause, and to automate the data collection and analysis process as much as possible. The author feels that the objectives were met through the various studies carried out in this thesis.

The research employed an empirical approach during all stages. Prototypes of each tool such as the TypingCollector, the CE questionnaire, and the categorisation method were built, and then tested with real participant data, the results of which directed further improvements. Since children's typing was a little studied area, this approach was particularly suited as it allowed for the adaptation of these tools for use with young children.

11.1 CONTRIBUTION OF THE THESIS

The thesis is that *'there is a set of typing error behaviours that are specific to children in phrase-copying typing'*. Further major contributions of this thesis come from the work carried out to enable the author to support this thesis.

In summary, a comparable typing data collection method for children and adults was designed (Chapter 5 and Chapter 6) along with a method of collecting their computer experience (Chapter 7). A new categorisation method was designed that encompassed all typing errors that children and adults made, without making any assumptions to their cause (Chapter 8). This categorisation method was then automated for more accurate, consistent and faster analysis

(Chapter 9). Finally, a large typing sample collected from children and adults provide empirical support for the thesis (Chapter 10).

11.1.1 Comparable Typing Data Collection Method

The first objective addressed to answer the main research question was designing a typing data collection method that caused as little bias as possible to the results. This meant that the method must cause minimum difficulties for the youngest participants.

Previous studies on text input with children used data collection methods commonly used with adults (Roussos, 1992; Read et al., 2001). Although suitable for adults, it was not known whether they were suitable for use with children. Presenting the text to be typed on paper and asking the participant to copy type onto a computer screen was popular with many typing studies with adults. However, Chapter 6 showed that this method caused several issues when used with young participants. Instead, a screen-to-screen approach, where both PT and TT are displayed on the same screen, in the same font size and style, was chosen.

Existing phrase sets to show the participants (Kucera and Francis, 1967; James and Reischel, 2001; MacKenzie and Soukoreff, 2003) contained words unsuitable for use with children. A new phrase set was designed that only contained words suitable for children aged six years and upwards (Chapter 5). This ensured that the adults did not have an unfair linguistic advantage over the children.

Although it was not possible to control the participants' previous computer experiences (CE), it was important to be able to measure their CE in some way so that the range of CE could be considered. To do this, Computer and Typing Experience Questionnaires (CTEQs) were developed, one for children and another for adults. The CTEQs extended existing CE questionnaires (Weil et al., 1990; Kinzie et al., 1994; Igbaria et al., 1995) by updating the tasks referenced (such as Facebook and Flickr), and ask specific questions about typing that previous questionnaires did not ask. Two Visual Analogue Scales developed to help children in answering CE questions, were found to be successful with even the youngest participants of this research. The study found that children were able to answer questions regarding the length and frequency of computer use as well as report on their perception of how good they are at typing. The Thumbs-Up scale and Frequency of Use scale can be used to ask questions regarding any task. They are good additional tools in toolkits designed to gather data from children, such as the Fun Toolkit (Read et al., 2001). The third study in Chapter 7 also found that children were able to give consistent answers between paper-based and computer-based CTEQ, suggesting that the two forms of administrations can be used interchangeably.

To provide a consistent data collection environment throughout the several schools visited for data collection, The TypingCollector was developed. Whereas previous data collecting software (Soukoreff and MacKenzie, 2003; Wobbrock and Myers, 2006) only focused on the typing task, the TypingCollector carried out the data collection process from start to finish (demographic questions, CE questions, selection of phrases to show, displaying of phrase and recording of the typing) with minimum interventions from the researcher. This enabled the researcher to carry out data collection with a whole classroom of children at a time.

11.1.2 The Categorisation Method

One of the major contributions of this thesis is the new categorisation method. Existing typing error categorisation methods were based on typing errors collected from adult typists. Using such methods to categorise typing errors would fail to detect any errors that may be unique to children. In addition, many methods omitted some typing errors, whereas others broke larger typing errors into numerous smaller errors.

A new categorisation method was defined based on those error types that combined real typing errors made by children and typing errors found in literature. The categorisation method was careful not to make assumptions as to the causes of these errors. This was important, since it was unknown as to whether the theoretical explanations found in literature of these error types in adults would apply to young children. The new categorisation method defined error types based on observable factors. It also removed typing errors that assumed formal typing training such as homologous errors (Gentner et al., 1983), since many people now have their own idiosyncratic method for typing that does not conform to this assumption.

The categorisation method was translated into a Java program that carried out the categorisation process. It took the work carried out by Soukoreff and MacKenzie (2001) and Wobbrock and Myers (2006) on character-level analysis of typing errors, and extended it to include further character-level errors as well as word-level errors. This allowed for an efficient and consistent categorisation, and a methodical reduction of ambiguities. The automation of the incredibly time consuming categorisation of typing errors means that larger studies with more data can be carried out with less resources.

11.1.3 Comparing Typing Errors of Children and Adults

Using the TypingCollector, a large corpus of typing data was collected from 231 children and 229 students. Their typing errors were detected and categorised using the TypingAnalyser.

The study found that there is indeed a difference between the ways that children and adults make typing mistakes. A simple analysis of typing error rates alone confirmed that there were significant differences between the amount of typing errors and corrections made by the two groups.

Another major contribution of this thesis lies in the investigation of each typing error category. This revealed rich information as to how the children and adults differed. It found a group of typing errors that were highly specific to the children's group. Many of these errors related to the misunderstanding of the rules of the copying task by the children. Adults rarely made these error types. A second group of typing errors were found to be more child-prone, where adults made some of these errors, but children made many more. These error types mostly related to errors that occur during the typing task.

These findings suggest that young children have a different process of typing to adults. In light of this, cognitive models constructed around how adults carry out typing cannot be directly applied to children. Each aspect of our understanding on how adults type must be tested with children before they can be said to also apply to children.

The decision to use error types instead of error rates was indeed the best choice for answering the main research question. In Chapter 10, it was shown that although error rates were able to show that there were significant differences between children and adults, it could not provide any details of how the differences occurred. The analysis of the error types themselves gave an additional dimension to the answer of 'yes, they are different', by being able to highlight a set of error types particular to children and another for the adults.

11.1.4 Caveats

In order to complete a coherent narrative to this thesis, restrictions were placed upon its scope. There were several factors that were not included in the studies reported, but are nevertheless highly interesting issues that should be studied further. Investigating these factors would also increase both the internal and external validity of the thesis of this work.

In Chapter 5 the Children's Phrase Set (CP Set) was manually selected from children's books following minimum restrictions, such as no symbols and no capital letters with the exception of the words 'I', 'June' and 'July'. The CP Set increased the internal validity of the final comparison study between children and adults' typing (Chapter 10), by reducing words with which children were likely to be unfamilar. However, much more rigour should be applied in the initial selection process of the phrases. A large body of writing suitable for young children should be used as an input for a selection algorithm based on a set of rules. Although the maximum word count of seven words per phrase was reasonable, minimum word count of three seems too low to provide useful typing data for a phrase-copying task. The minimum word count therefore should be set higher, perhaps at five words per phrase, or even for all phrases to have exactly the same number of words. Similarly, there should be a maximum limit on the number of letters in a word to filter out very long words. Further, grammatical rules such as removing all proper nouns should also be applied to the phrases.

In studies evaluating memory in both children and adults, concrete words have been observed as being considerably easier to remember than abstract words (Paivio and Begg, 1971). Vellutino and Scanlon (1985) also found that children who were poor readers had more difficulties in recalling abstract words than normal readers. Therefore researchers should take this into consideration when designing a phrase set. Another interesting choice of phrase selection would be to use nonsense words (words that are made up of random selection of letters such as those used in Shaffer and Harwick (1969) instead. This would ensure that none of the participants would have the advantage of being more familiar with the language used in the presented text.

There are further considerations that should be made regarding the abilities of the participants. Whether the participant is a good or poor reader, whether they are visual or non-visual thinkers, whether they suffer from learning disabilities that affect their short-term memory (such as dyslexia) all have an effect on how well they will read and remember the phrases shown to them.

Since the investigation into the relationship between computer experience (Chapter 7) and typing performance was outside the scope of the thesis, this was not examined in Chapter 10. However, such study would be uesful in reducing the 19 questions Adult CTEQ to those that most closely relate to typing performance. In contrast, the children's CTEQ only contained six questions covering five aspects of computer experience. It is possible that an aspect that was omitted from the Children's CTEQ may relate more closely to typing performance.

In addition to the keys pressed during the phrase-copy typing task, the TypingCollector gathered timings of these key presses. The current work has only used these timing data to classifiy simultaneous key presses such as NT-Mu and CT-Mu. However, the timing data is a rich source of typing behaviour information that should be used to assist in understanding typing behaviours.

The time taken between each key press (inter-keypress times) has been used as a measure of typing performance in adult typing in many studies (see Section 2.2.2). A closer examination of the inter-keypress timings surrounding a typing error could reveal further details of how specific typing errors occur. It could also be used as a indicator of possible typing errors such as the study carried out by Shaffer (1975) in which they found that many omission errors were followed by a keystroke that took twice as long as the overall median inter-keypress timing.

The overall time taken for the participant to type each phrase is available from the TypingCollector log. The variation in the time taken for each participant group could be a useful additional dimension in the typing profile of each participant. This would be used to establish whether the typist is more likely to be a child than an adult. However, care should be taken to not use timing alone in establishing the participant's age group.

Since the focus of this thesis was in the analysis of the typing errors made by the participants, the error rates themselves are only briefly reported. It is already known that the design of the Presented Text influences typing speed in adults (Salthouse, 1986). There is an interesting question as to whether the word length of the Presented Text or how the frequency of each word appears in English relate to the participants' error rates.

11.2 APPLICABILITY OF THE FINDINGS AND FUTURE WORK

The study of adult CE questions (Chapter 7) showed that by breaking down CE into much smaller aspects, and asking questions on each of them, allowed for clearly identifying the effects of the sampling method had on the sample's CE. Another use for this questionnaire would be to investigate which questions in the CE correlate most highly with the participant's typing. Since the adult CE questionnaire contained 19 questions, there is also scope for narrowing down this questionnaire to only those most relevant indicators of typing performance.

Additionally, children and adults currently have individual CE questionnaires. This was due to the children having difficulties answering certain CE questions that referred to particular software names. The child CE questionnaire was also restricted to a low number of questions. Due to the separation of the two questionnaires, it is currently not possible to place the two participant groups on a continuous scale of CE. It would be more useful in a study of CE and typing performance to have one questionnaire to measure children and adult CE in an appropriate manner.

The TypingAnalyser can classify letter-level and word-level errors. The TypingAnalyser should be extended to classify phrase level errors. In Section 10.3.5.2, 10.3.6.4 and 10.3.8.2, it was found that some error types (such as consistent Duplicated Spaces, Omitted Spaces and typing all the letters in uppercase) inflated the total number of errors. The TypingAnalyser will be able to produce even more accurate information if it was able to detect consistent behaviours and count them as one error, rather than as many errors.

One of the limitations of this research was that the findings were based on a relatively small sample size, selected from one particular region in the UK. The child participants were all recruited from local state primary schools. The adult participants were recruited from the author's university department. These two sample groups were chosen due to ease of access to them. However, it is acknowledged that the ease of access was costly in term of the representativeness of the sample. There were no child participants taken from private schools or from any other part of the country. The adult sample were mostly males in their 20s, likely to have higher than average computer skills. Generalisation of the findings to the larger population is therefore limited at this point.

The groups of typing errors specific to the two participant groups indicate that there are notable differences in the way children and adults make typing mistakes. However, some children made no child-specific or child-prone errors, but made some adult-prone errors. Similarly, some adult participants made child-prone errors. It is not possible at this stage to pick one person's typing and state clearly whether the participant is a child or an adult - only that they are more likely to be one than the other.

One of the main limitations of the findings is the sampling method used, in particular the adult sample. The use of first-year undergraduate computing students caused a skew in gender, age and CE that is not representative of the general adults population. A much larger study is required to increase the external validity of the findings of this thesis. Samples should be taken from different regions of the country and perhaps other English speaking countries such as America, Canada and Australia. A wider age, gender and computer skill range should be targeted in the adult sample. A larger study with wider range of participants will assist in confirming the child-specific, child-prone and adult-prone error types found in this study. It may even reveal new error types yet to be defined in the categorisation method.

Another point of inquiry is why the observed differences in the way children and adults make typing mistakes occur in the first place. So far, this research has focused on the detection of the differences between them. Analysis of the timing of key presses and other approaches discussed in Section 11.1.4 should be used to explain or eliminate theories about how a particular error type occurs. Investigations into each participant's psychological makeup such as cognitive and motor-control abilities, reading abilities and whether they are visual or non-visual learners may reveal psychological reasons behind one group being more prone to an error type than the other.

Lastly, once more studies are carried out to establish a more clear set of typing behaviours that are particular to children or to adults, the differences in their typing could be used beyond simply comparing between the two groups. For example, a chat-room, forum or a website that is designed for children could analyse all the users' typing. If a user, posing as a child makes a large amount of adult-specific errors and very few child-specific errors, then a warning could be sent to the administrators that this user may be an adult pretending to be a child.

11.3 Concluding Remarks

The studies in this suggest that children are not the same as adults in the way they type. Therefore, application of theoretical models based on how adults type cannot be applied in a straightforward manner to children. Further studies investigating the causes of children's typing are required before mental models of their typing behaviour can be constructed.

This thesis focused on the differences between children and adult's typing errors. Other works have compared typing skills (Grudin, 1983a), and age difference (Salthouse, 1984). However, differences may exist between other participant groups. How does being a native English speaker differ from a non-native English speaker? Is it possible that dyslexic typists differ from non-dyslexic typists? Is it possible to detect that someone is dyslexic from the way they make typing mistakes? The tools created in this thesis can be used in gathering and analysing typing data to answer these questions.

12 BIBLIOGRAPHY

Aera and Ncm (1999). American Educational Research Association and National Council on Measurement- Standards for Educational and Psychological Testing. Washington DC, American Psychological Association.

Arch, E. C. and Cummins, D. E. (1989). 'Structured and Unstructured Exposure to Computers: Sex Differences in Attitude and Use among College Students.' Sex Roles **20**(5-6): 245-254.

Bandura, A. (1977). 'Self-Efficacy: Toward a Unifying Theory of Behavior Change.' Psycholgy Review **84**: 191-215.

Bandura, A. and Adams, N. E. (1977). 'Analysis of Self-Efficacy Theory of Behavioral Change.' Cognitive Therapy and Research **1**: 287-310.

Bandura, A., Adams, N. E. and Beyer, J. C. (1977). 'Cognitive Processes Mediating Behavioral Change.' Journal of Personality and Social Psychology **35**: 125-139.

Banister, P., Burman, E., Parkey, I., Taylour, M. and Tindall, C. (1998). Qualitative Methods in Psychology - a Research Guide. UK, Open University Press.

Beckers, J. J. and Schmidt, H. G. (2003). 'Computer Experience and Computer Anxiety.' Computers in Human Behavior **19**(6): 785-797.

Bell, A. (2007). 'Designing and Testing Questionnaires for Children.' Journal of Research in Nursing **12**(5): 461-469.

Bix, L. (2002). 'The Elements of Text and Message Design and Their Impact on Message Legibility: A Literature Review.' Journal of Design Communication(4).

Blandford, A., Cox, A. L. and Cairns, P. (2008). Controlled Experiments. Research Methods for Human-Computer Interaction. Cairns, P. and Cox, A. L. Cambridge, UK, Cambridge University Press: 1-16.

Book, W. F. (1925). Learning to Typewrite. With a Discussion of the Psychology and Pedagogy of Skill, The Gregg Publishing Company.

Borgers, N., Hox, J. and Sikkel, D. (2004). 'Response Effects in Surveys on Children and Adolescents: The Effect of Number of Response Options, Negative Wording, and Neutral Mid-Point.' Quality and Quantity **38**(1): 17-33.

Bryman, A. (2004). Social Research Methods (2nd Ed.). Oxford, Oxford University Press Inc.

Buchanan, T., Johnson, J. A. and Goldberg, L. R. (2005). 'Implementing a Five-Factor Personality Inventory for Use on the Internet.' European Journal of Psychological Assessment **21**: 115-127.

Burke, J., Burke, K. and Baker, J. (1995). 'Test Retest Reliability in Psychiatric Patients of the Sf-36 Health Survey.' International Journal of Methods in Psychiatric Research **5**: 189-194.

Busch, T. (1995). 'Gender Differences in Self-Efficacy and Attitudes toward Computers.' Journal of Educational Computing Research **12**(2): 147-158.

Butsch, R. L. C. (1932). 'Eye Movements and the Eye-Hand Span in Typewriting.' Journal of Educational Psychology **23**(2): 104-121.

Butts, L. and Cockburn, A. (2001). An Evaluation of Mobile Phone Text Input Methods. Third Australasian User Interface Conference, Melbourne, Australia.

Buzing, P. (2003) Comparing Different Keyboard Layouts - Aspects of Qwerty, Dvorak and Alphabetical Keyboards.

Chua, S. L., Chen, D. T. and Wong, A. F. L. (1999). 'Computer Anxiety and Its Correlates: A Meta-Analysis.' Computers in Human Behavior **15**: 609-623.

Clem, J. (1929). The Technique of Teaching Typewriting. New York, The Gregg Publishing Company.

Colley, A. M., Gale, M. T. and Harris, T. A. (1994). 'Effects of Gender Role Identity and Experience on Computer Attitude Components.' Journal of Educational Computing Research **10**: 129-137.

Cook, T. D. and Campbell, D. T. (1979). Quasi-Experimentation: Design and Analysis for Field Settings. Boston, Houghton Mifflin.

Cooling, Ed. (1998). Read Me a Story, Please, Orion Children's Books, UK.

Cooper, W. E. (1983a). Chapter 1 - Introduction. Cognitive Aspects of Skilled Typewriting. Cooper, W. E. New York, Springer-Verlag.

Cooper, W. E., Ed. (1983b). Cognitive Aspects of Skilled Typewriting, Springer-Verlag, New York.

Cronbach, L. J. (1990). Essentials of Psychological Testing (5th Ed.). New York, Harper Collins.

Dambrot, F. H., Watkins-Malek, M. A., Silling, S. M., Marshall, R. S. and Garver, J. (1985). 'Correlates of Sex Differences in Attitudes toward and Involvement with Computers.' Journal of Vocational Behaviour **27**: 71-86.

Davis, D. D. W. (1935) An Analysis of Student Errors on the Universal and on the Dvorak-Dealey Simplified Typewriter Keyboard. Master of Arts. University of Washington,

Davis, F. D., Bagozzie, R. P. and Warshaw, P. R. (1989). 'User Acceptance of Computer Technology: A Comparison of Two Theoretical Models.' Management Science **35**: 982-1003.

De Lange, R. W., Esterhuizen, H. L. and Beatty, D. (1993). 'Performance Differences between Times and Helvetica in a Reading Task.' Electronic Publishing **6**(3): 241-248.

Delcourt, M. A. B. and Kinzie, M. B. (1990). Attitude toward Computer Technologies (Act) and Self-Efficacy for Computer Technologies (Sct) McGill University, Department of Educational Psychology and Counselling, Montreal, Quebec, Canada.

Delcourt, M. A. B. and Kinzie, M. B. (1993). 'Computer Technologies in Teacher Education: The Measurement of Attitudes and Self-Efficacy.' Journal of Research & Development in Education **27**: 31-37.

Diehl, M. J. and Seibel, R. (1962). 'The Relative Importance of Visual and Auditory Feedback in Speed Typewriting.' Journal of Applied Psychology **46**(5): 365-369.

Does (1989). Department of Education and Science - Information Technology from 5 to 16, Her Majesty's Stationery Office, London.

Dvorak, A., Merrick, N. L., Dealey, W. L. and Ford, G. C. (1936). Typewriting Behaviour, American Book Company, USA.

Eagly, A. H. and Chaiken, S. (1993). The Psychology of Attitudes. Fort Worth, Harcourt Brace Jovanovich.

Ertmer, P. A., Evenbeck, E., Cennamo, K. S. and Lehman, J. D. (1994). 'Enhancing Self-Effcacy for Computer Technologies through the Use of Positive Classroom Experiences ' Educational Technology, Research and Development **42**: 45-62.

Farthing, G. W. (1992). The Psychology of Consciousness. Englewood Cliffs, NJ, Prentice Hall.

Fendrick, P. (1937). 'Hierachical Skills in Typewriting.' Journal of Educational Psychology **28**(8): 609-620.

Ford, G. C. (1928) A Study of Typewriting Errors. Master of Arts. University of Washington,

Fox, J. G. and Stansfield, R. G. (1964). 'Digram Keying Times for Typists.' Ergonomics **7**: 317-320.

Gentner, D. R. (1982). 'Evidence against a Central Control Model of Timing in Typing.' Journal of Experimental Psychology: Human Perception and Performance **8**(6): 793-810.

Gentner, D. R. (1983). 'The Acqusition of Typewriting Skill.' Acta Psychologica **54**: 233-248.

Gentner, D. R., Grudin, J. T. and Conway, E. (1980). Finger Movements in Transcription Typing, Technical Report 8001. San Diego, CA, Center for Human Infromation Processing, University of California.

Gentner, D. R., Grudin, J. T., Larochelle, S., Norman, D. A. and Rumelhart, D. E. (1983). A Glossary of Terms Including a Classification of Typing Errors. Cognitive Aspects of Skilled Typewriting. Cooper, W. E., Springer Verlag**:** 39-43.

Gentner, D. R. and Larochelle, S. (1988). 'Lexical, Sublexial, and Peripheral Effects in Skilled Typewriting.' Cognitive Psychology **20**: 524-548.

Gilroy, F. D. and Desai, H. B. (1986). 'Computer Anxiety: Sex, Race, and Age.' International Journal of Man-Machine Studies **25**: 711-719.

Grudin, J. T. (1983a). Error Patterns in Novice and Skilled Transcription Typing. Cognitive Aspects of Skilled Typing. Cooper, W. E. New York, USA, Springer-Verlag**:** 121-139.

Grudin, J. T. (1983b). 'Non-Hierarchic Specification of Components in Transcription Typewriting.' Acta Psychologica **54**: 249-262.

Hallfors, D., Kathapoush, S., Kadushin, C., Watson, K. and Saxe, L. (2000). 'A Comparison of Paper Vs. Computer-Assisted Self Interview for School Alcohol, Tobacco, and Other Drug Surveys.' Evaluation and Program Planning **23**(2): 149-155.

Hayes, V. (1978). 'Relationship between Typewriting Rate and Responses to Individual Characters on the Typewriter Keyboard.' Perceptual and Motor Skills **46**: 1339-1343.

Hayes, V. and Reeve, T. G. (1980). 'Role of Visual Feedback for Response Guidance and Response Confirmation in Typewriting.' Perceptual and Motor Skills **50**: 1047-1056.

Hayes, V., Wilson, G. D. and Schaffer, R. L. (1977). 'Typewriting Rate as a Function of Reaction Time.' Perceptual and Motor Skills **45**: 1179-1184.

Heinssen, R. K., Glass, C. R. and Knight, L. A. (1987). 'Assessing Computer Anxiety: Development and Validation of the Computer Anxiety Rating Scale.' Computers in Human Behaviour **3**(49-59).

Henderson, R., Deane, F., Barrelle, K. and Mahar, D. (1995). 'Computer Anxiety: Correlates, Norms and Problem Definition in Health Care and Banking Employees Using the Computer Attitude Scale.' Interacting with Computers **7**(181-193).

Hershman, R. L. and Hillix, W. A. (1965). 'Data Processing in Typing: Typing Rate as a Function on Kind of Material and Amount Exposed.' Human Factors **7**(5): 483-492.

Hill, T., Smith, N. D. and Mann, M. F. (1987). 'Role of Efficacy Expectations in Predicting the Decision to Use Advanced Technologies: The Case of Computers.' Journal of Applied Psychology

72(2): 307-313.

Hiraga, Y., Ono, Y. and Hisao, Y. (1980). An Analysis of the Standard English Keyboard. 8th Conference on Computational Linguistics, Tokyo, Japan.

Hoke, R. E. (1922). The Improvement Speed and Accuracy in Typewritng, John Hopkins Press.

Hopkins, C. (1989). Beyond the Qwerty Keyboard. Computers and the Primary Curriculum 3-13. Crompton, R. London, UK, Falmer Press.

Hornby, A. S. (2007). Oxford Advanced Learner's Dictionary, Seventh Edition with Oxford 3000 Vocabulary Trainer, Oxford University Press.

Howard, G. S. and Smith, R. D. (1986). 'Computer Anxiety in Management: Myth or Reality.' Communications of ACM **29**: 611-615.

Igbaria, M. and Chakrabarti, A. (1990). 'Computer Anxiety and Attitudes Towards Microcomputer Use.' Behaviour and Information Technology **9**(3): 229-241.

Igbaria, M., Iivari, J. and Maragahh, H. (1995). 'Why Do Individuals Use Computer Technology? A Finnish Case Study.' Information and Management **29**: 227-238.

Inhoff, A. W. (1991). 'Word Frequency During Copytyping.' Journal of Experimental Psychology: Human Perception and Performance **17**(2): 478-487.

Isokoski, P. and Raisamo, R. (2000). Device Independent Text Input: A Rationale and an Example. AVI 2000, Palermo, Italy. ACM.

James, C. L. and Reischel, K. M. (2001). Text Input for Mobile Devices: Comparing Model Prediction to Actual Performance. Proceedings of CHI2001. ACM, New York.

Janelle, R. V. (1974). Eastern Region Trial of the Dvorak Simplified Keyboard. Eastern Region, USA, American Telegraph and Telephone, Western Electric.

Jones, T. and Clarke, V. A. (1995). 'Diversity as a Determinant of Attitudes: A Possible Explanation of the Apparent Advantage of Single-Sex Setting.' Journal of Educational Computing Research **12**: 51-64.

Jordan, E. W. and Stroup, D. F. (1982). 'The Behavioral Antecedents of Computer Fear.' Journal of Data Education **22**: 7-8.

Joyce, B. G. and Moxley, R. A. (1989). Comparing Children's Typing Skills Using the Dvorak and Qwerty Keyboards on a Microcomputer.

Kano, A. (2005). Evaluating Phrase Sets for Use with Text Entry Method Evaluation with Dyslexic Participants. Workshop on Improving and Assessing Pen-Based Input Techniques at HCI 2005, Edinburgh, UK.

Kano, A., Horton, M. and Read, J. C. (2010). Thumbs-up Scale and Frequency of Use Scale for Use in Self Reporting of Children's Computer Experience. NordiCHI 2010, Reykjavik, Iceland. ACM.

Kano, A., Read, J. C. and Dix, A. (2006). Children's Phrase Set for Text Input Method Evaluations. Proceedings of the Fourth Nordic Conference on Human-Computer Interaction - NordiCHI, Oslo, Norway. ACM Press.

Kano, A., Read, J. C., Dix, A. and Mackenzie, I. S. (2007). Expect: An Expanded Error Categorisation Method for Text Input. HCI 2007, Lancaster, UK. ACM.

Kay, R. H. (1990). 'Predicting Student Teacher Commitment to the Use of Computer.' Journal of Educational Computing Research **6**: 299-309.

Kay, R. H. (1992). 'An Analysis of Methods Used to Examine Gender Differences in Computer Related Behaviour.' Journal of Educational Computing Research **8**: 323-336.

Kinzie, M. B., Delcourt, M. A. B. and Powers, S. M. (1994). 'Computer Technologies: Attitudes and Self-Efficacy across Undergraduate Disciplines.' Research in Higher Education **35**(6): 745-768.

Kirkman, C. (1993). 'Computer Experience and Attitudes of 12-Year-Old Students: Implications for the Uk National Curriculum.' Journal of Computer Assisted Learning **9**: 51-62.

Koslowsky, M., Hoffman, M. and Lazar, A. (1990). 'Predicting Behaviour on a Computer from Intentions, Attitude, and Experience.' Current Psychology: Research and Review **9**(1): 75-83.

Krevolin, N. (1965). 'How Can We Best Implement Elementary School Typing Courses?' Business Education World **46**(3): 11-18.

Kucera, H. and Francis, W. N. (1967). Computational Analysis of Present-Day American English. Providence, R. I., Brown University Press.

Larochelle, S. (1984). 'Some Aspects of Movements in Skilled Typewriting.' Attention and Performance. Control of Language Processes **10**: 43-54.

Lashley, K. S. (1951). The Problem of Serial Order in Behavior. Cerebral Mechanisms in Behavior. Jefferess.

Lee, J. A. (1986). 'The Effects of Past Computer Experience on Computerized Aptitude Test Performance.' Educational and Psychological Measurement **46**(3): 727-733.

Lessenberry, D. D. (1928). Error Analysis Chart, L. C. Smith and Corona Typewriters, Inc. Syracuse, N.Y.

Levenshtein, V. I. (1966). 'Binary Codes Capable of Correcting Deletions, Insertions and Reversals.' Sovient Physics-Doklandy **10**: 707-710.

Lewis, J. R. (1999). Input Rates and User Preference for Three Small-Screen Input. Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, Santa Monica, CA. Human Factors and Ergonomics Society.

Loeber, R. and Farrington, D. (1989). Development of a New Measure of Self-Reported Antisocial Behavior for Young Children: Prevalence and Reliability. Cross-National Research in Self-reported Crime and Delinquency. Klein, M. Dordrecht, Kluwer Academic.

Logan, F. A. (1999). 'Errors in Copy Typewriting.' Journal of Experimental Psychology: Human Perception and Performance **25**(6): 1760-1773.

Logan, G. D. (1982). 'On the Ability to Inhibit Complex Movements: A Stop-Signal Study of Typewriting.' Journal of Experimental Psychology: Human Perception and Performance **8**(6): 778-792.

Loyd, B. and Fressard, C. (1984). The Effects of Sex, Age, and Computer Experience on Computer Attitudes. Annual Meeting of the Eastern Educational Research Association of West Palm Beach, FL.

Lukin, M., Dowd, E. and Plake, B. (1985). 'Comparing Computerized Versus Traditional Psychological Assessment.' Computers in Human Behavior **1**: 49-58.

Lund, O. (1999) Knowledge Construction in Typography: The Case of Legibility Research and the Legibility of Sans Serif Typefaces. PhD Thesis. Department of Typography & Graphic Communication, University of Reading, Reading.

Mackay, W. E. and Fayard, A.-L. (1997). Hci, Natural Science and Design: A Framework for Triangulation across Disciplines. DIS '97 Proceedings of the 2nd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques. ACM, New York.

Mackenzie, I. S. (2002a). Kspc (Keystrokes Per Character) as a Characteristic of Text Entry Techniques. Proceedings of the Fourth International Symposium on Human Computer Interaction with Mobile Devices, Heidelberg, Germany. Springer-Verlag.

Mackenzie, I. S. (2002b). Mobile Text Entry Using Three Keys. Proceedings of the Second Nordic Conference on Human-Computer Interaction - NordiCHI 2002. ACM Press.

Mackenzie, I. S. (2006). Teps, a Shortened Name for the Phrase Set, Personal Communication.

Mackenzie, I. S. (2007). Evaluation of Text Entry Techniques. Text Entry Systems: Mobility, Accessibility, Universality. MacKenzie, I. S. and Tanaka-Ishii, K. San Francisco, Morgan Kaufmann Publishers.

Mackenzie, I. S., Nonnecke, B., Mcqueen, C., Riddersma, S. and Meltz, M. (1994). A Comparison of Three Methods of Character Entry on Pen-Based Computers. Human Factors and Ergonomics Society 38th Annual Meeting, Santa Monica, CA. Human Factors Society.

Mackenzie, I. S. and Soukoreff, R. W. (2002a). A Character-Level Error Analysis Technique for Evaluating Text Entry Methods. Proceedings of the Second Nordic Conference on Human-Computer Interaction, Aarhus, Denmark. ACM Press

Mackenzie, I. S. and Soukoreff, R. W. (2002b). Text Entry for Mobile Computing: Models and Methods, Theory and Practice. Human-Computer Interaction. Lawrence Earlbaum Associates, Inc.

Mackenzie, I. S. and Soukoreff, R. W. (2003). Phrase Sets for Evaluating Text Entry Techniques. Extended Abstracts of the ACM Conference on Human Factors in Computing Systems - CHI 2003. ACM Press.

Macneilage, P. F. (1964). 'Typing Errors as Clues to Serial Ordering Mechanisms in Language Behaviour.' Language and Speech **7**: 1144-1159.

Markopoulos, P., Read, J. C., Macfarlane, S. and Hoysniemi, J. (2008). Evaluating Children’s Interactive Products. Burlington, MA, USA, Morgan Kaufmann.

Martin, D. W. (2008). Doing Psycology Experiments (7th Ed.). Belmont, CA, Thomson Wadsworth.

Mayzner, M. S. and Tresselt, M. E. (1965). 'Table of Single Letter and Digram Frequency Counts for Various Word-Length and Letter-Position Combinations.' Psychonomic Monograph Supplements 1: 13-32.

Millistein, S. G. (1987). 'Acceptability and Reliability of Sensitive Information Collected Via Computer Interview.' Educational and Psychological Measurement **47**(2): 523-533.

Murphy, C. A., Coover, D. and Owen, S. (1988). Assessment of Computer Self-Efficacy: Instrument Development and Validation. the annual meeting of the National Council on Measurement in Education, San Diego.

Naus, M. J., Philipp, L. M. and Samsi, M. (2009). 'From Paper to Pixes: A Comparison of Paper and Computer Formats in Psychological Assessment.' Computers in Human Behavior **25**: 1-7.

Nelson, L. J., Wiese, G. M. and Cooper, J. (1991). 'Getting Started with Computers: Experience, Anxiety and Relational Style.' Computers in Human Behavior **7**: 185-202.

Newman, D. L. and Clure, G. (1984). Computer Anxiety in Elementary School Children: The Development of an Instrument and Its Correlates. The Meeting of the Rocky Mountain Research Association, Oklahoma City, USA.

Norman, D. A. and Rumelhard, D. E. (1983). Studies of Typing from the Lnr Research Group. Cognitive Aspects of Skilled Typewriting. Cooper, W. E. New York, Springer-Verlag.

Noyes, J. (1983). 'The Qwerty Keyboard, a Review.' International Journal of Man-Machine Studies **18**: 265-281.

O'shea, T. and Shelf, J. (1983). Learning and Teaching with Computers: Artificial Intelligence in Education. Sussex, UK, Harvester Press.

Okebukola, P. A., Sumampouw, W. and Jegede, O. J. (1992). 'The Experience Factor in Computer Anxiety and Interest.' Journal of Educational Technology Systems **20**: 221-229.

Oniszczak, A. and Mackenzie, I. S. (2004). A Comparison of Two Input Methods for Keypads on Mobile Devices. NordiCHI 2004, Tampere, Finland. ACM.

Opfer, D. S. (1932) Critical Study of the Typing Errors of Beginning Pupils in Four Los Angeles Senior High School. MA Thesis. University of Southern California, CA.

Paivio, A. and Begg, I. (1971). 'Imagery and Associative Overlap in Short-Term Memory.' Journal of Experimental Psychology **89**(1): 40-45.

Petitt, F. A. (2002). 'A Comparison of World-Wide Web and Paper-and-Pencil Personality Questionnaires.' Behavior Research Methods, Instruments & Computers **34**(1): 50-54.

Pouwer, F., Snoek, F. J., Van Der Ploeg, H. M., Heine, R. J. and Brand, A. N. (1998). 'A Comparison of the Standard and the Computerized Versions of the Well-Being Questionnaire (Wbq) and the Diabetes Treatment Satisfaction Questionnaire (Dtsq).' Quality of Life Research 7(1): 33-38.

Provins, K. A. and Glencross, D. J. (1968). 'Handwriting, Typewriting and Handedness.' Quarterly Journal of Experimental Psychology **20**(3): 282-289.

Rammstedt, B., Holzinger, B. and Rammsayer, T. (2004). 'Comparison of Paper-and-Pencil Vs. Computerised Administration of the Neo-Five Factor Inventory (Neo-Ffi).' Diagnostica **50**: 88-97.

Read, J. C. (2005). The Usability of Digital Ink Technologies for Children and Teenagers. HCI2005, Edinburgh, Scotland. Springer.

Read, J. C. and Horton, M. (2006). When Teenagers Ttype. British HCI 2006, London, UK.

Read, J. C., Macfarlane, S. J. and Casey, C. (2001). Measuring the Usability of Text Input Methods for Children. HCI2001, Lille, France. Springer-Verlag.

Read, J. C., Macfarlane, S. J. and Casey, C. (2002). Pens Behaving Badly - Usability of Pens and Graphics Tablets for Text Entry with Children. ACM UIST'02, Paris, France.

Robertson, S. I., Calder, J., Fung, P., Jones, A. and O'shea, T. (1995). 'Computer Attitudes in an English Secondary School.' Computer Education **24**(2): 73-81.

Rosen, L. D., Sears, D. C. and Weil, M. M. (1987). 'Computerphobia.' Behavior Research Methods, Instruments & Computers **19**: 167-179.

Rosen, L. D. and Weil, M. M. (1995). 'Computer Availability, Computer Experience and Technophobia among Public School Teachers.' Computers in Human Behavior **11**(1): 9-31.

Roussos, P. (1992) Effects of Keyboard Layout on Children's Performance and Interaction with Computers. Doctor of Philosophy. Department of Psychology, University of Leeds,

Rowe, C. E. (1931). 'Correcting Non-Letter Errors in Typing.' Journal of Business Education **V**: 30-31.

Rowe, J. L. (1959). 'Readin', Typin', and 'Rithmetic.' Business Education World **39**(5): 9-12.

Rozell, E. J. and Gardner, W. L. (1999). 'Computer-Related Success and Failure: A Longitudinal Field Study of the Factors Influencing Computer-Related Performance.' Computers in Human Behavior **15**(1): 1-10.

Russon, A. R. and Wanous, S. J. (1973). Philosophy and Psychology of Teaching Typewriting 2nd Ed. Cincinnati, Ohio, South-Western Publishing Co.

Salthouse, T. A. (1984). 'Effects of Age and Skill in Typing.' Journal of Experimental Psychology: General **113**(3): 345-371.

Salthouse, T. A. (1985). 'Anticipatory Processing in Transcription Typing.' Journal of Applied Psychology **70**(2): 264-271.

Salthouse, T. A. (1986). 'Perceptual, Cognitive, and Motoric Aspects of Transcription Typing.' Psychological Bulletin **99**(3): 303-319.

Schiller, P. and Moore, T. (2004). Do You Know the Muffin Man? - an Essential Preschool Literacy Resource, Gryphone House, MD.

Schoenleber, L. (1932) A Procedure for Error Analysis in Secondary School Typewriting. MA. Department of Commerce, Graduate Collecge of the State University of Iowa,

Schunk, D. H. (1985). 'Self-Efficacy and Classroom Learning.' Psychology in the Schools **22**: 208-223.

Schunk, D. H. (1989). Self-Efficacy and Cognitive Skill Learning. Research on Motivation in Education: Vol. 3. Goals and Cognitions. C. Ames and R. Ames. San Diego, Academic Press.

Scott, J. (2000). Children as Respondents: The Challenge for Quantitative Methods. Research with Children – Perspectives and Practices. Christensen, P. and James, A. London, UK, Falmer Press**:** 98-119.

Shaffer, L. H. (1975). 'Control Processes in Typing.' Quarterly Journal of Experimental Psychology **27**: 419-432.

Shiue, Y.-M. (2002). 'The Effects of Cognitive Learning Style and Prior Computer Experience on Taiwanese College Students' Computer Self-Efficacy in Computer Literacy Courses.' Journal of Educational Technology Systems **31**(4): 393-409.

Siann, G., Macleod, H., Glissov, P. and Durndell, A. (1990). 'The Effect of Computer Use on Gender Differences in Attitudes to Computers.' Computers in Education **14**: 183-191.

Silfverberg, M. (2007). Historical Overview of Consumer Text Entry Techniques. Text Entry Systems: Mobility, Accessibility, Universality. San Francisco, Morgan Kaufmann Publishers.

Smith, B. L., Caputi, P., Crittenden, N., Jayasuriya, R. and Rawstorne, P. (1999). 'A Review of the Construct of Computer Experience.' Computers in Human Behavior **15**: 227-242.

Smith, H. H. (1932). Problems in Diagnostic Testing and Remedial Teaching as Applied to Typewriting. Second Yearbook, Commercial Education Association of New York City and Vicinity**:** 53-68.

Soukoreff, R. W. (2010) Quantifying Text Entry Performance. PhD Thesis. Department of Computer Science, York Unveristy, Toronto, Ontario, Canada.

Soukoreff, R. W. and Mackenzie, I. S. (2001). Measuring Errors in Text Entry Tasks: An Application of the Levenshtein String Distance Statistic. Extended Abstract of the ACM Conference on Human Factors in Computing Systems (CHI), Seattle, WA. ACM: New York.

Soukoreff, R. W. and Mackenzie, I. S. (2003). Metrics for Text Entry Research: An Evaluation of Msd and Kspc, and a New Unified Error Metric. Ft. Lauderdale, Florida, USA, ACM Press

Szajna, B. and Mackay, J. M. (1995). 'Predictors of Computer-User Training Environment: A Path Analytic Study.' International Journal of Human-Computer Interaction **7**(2): 167-185.

Tinker, M. A. and Paterson, D. G. (1932). 'Studies of Typographical Factors Influencing Speed of Reading: X. Style of Typeface.' Journal of Applied Psychology **16**(6): 605-613.

Todman, J. and Lawrenson, H. (1992). 'Computer Anxiety in Primary Schoolchildren and University Students.' British Educational Research Journal **18**(1): 63-72.

Todman, J. and Monaghan, E. (1994). 'Qualitative DiErences in Computer Experience, Computer Anxiety, and Students Use of Computers: A Path Model.' Computers in Human Behaviour **10**(4): 529-539.

Truman, J., Robinson, K., Evans, A. L., Smith, D., Cunningham, L., Millward, R. and Minnis, H. (2003). 'The Strengths and Difficulties Questionnaire - a Pilot Study of a New Computer Version of the Self-Report Scale.' European Child & Adolescent Psychiatry **12**: 9-14.

Vellutino, F. R. and Scanlon, D. M. (1985). 'Free Recall of Concrete and Abstract Words in Poor and Normal Readers.' Journal of Experimental Child Psychology **39**(2): 363-380.

Vereecken, C. A. and Maes, L. (2006). 'Comparison of a Computer-Administered and Paper-and-Pencil-Administered Questionnaire on Health and Lifestyle Behaviours.' Journal of Adolescent Health **38**: 426-432.

Walker, S. and Reynolds, L. (2002). 'Serifs, Sans Serifs and Infant Characters in Children's Reading Books.' Information Design Journal **11**(2(3)): 106-122.

Webb, P. M., Zimet, G. D., Fortenberry, J. D. and Blythe, M. J. (1999). 'Comparability of a Computer-Assisted Versus Written Method for Collecting Health Behavior Information from Adolescent Patients.' Journal of Adolescent Health **24**(6): 383-388.

Weil, M. M., Rosen, L. D. and Wugalter, S. E. (1990). 'The Etiology of Computerphobia.' Computers in Human Behaviour **6**: 361-379.

Wells, F. L. (1916). 'On the Psychomotor Mechanism of Typewriting.' American Journal of Psychology **XXXVII**: 47-70.

White, W. T. (1932). Typing for Accuracy, H. M. Rowe Company, Balitimore & Chicago, USA [Reprinted in Dvorak et al. Typewriting Behaviour, American Book Company, New York, 1936].

Wobbrock, J. O. (2007). Measures of Text Entry Performance. Text Entry Systems. MacKenzie, I. S. and Tanaka-Ishii, K. San Francisco, Morgan Kaufmann Publishers**:** 47-74.

Wobbrock, J. O. and Myers, B. A. (2006). Analyzing the Input Stream for Character- Level Errors in Unconstrained Text Entry Evaluations. ACM Transactions on Computer-Human Interaction (TOCHI). ACM.

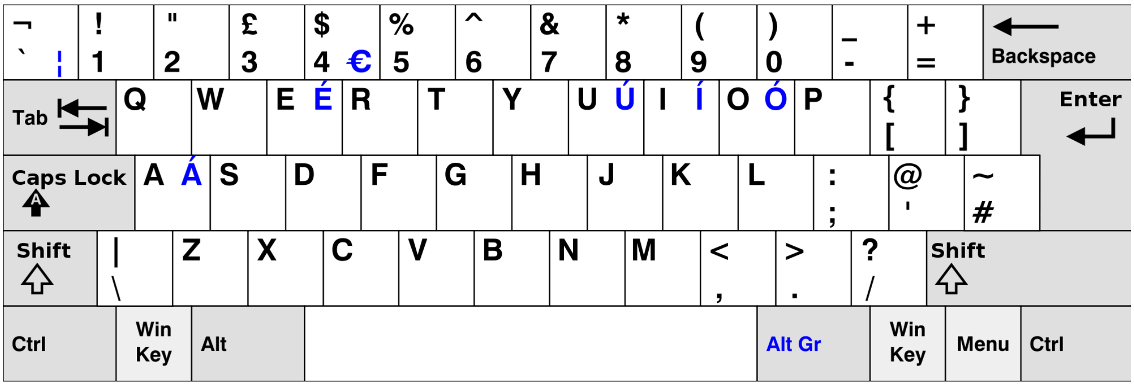
Wood, B. D. and Freeman, F. N. (1932). An Experimental Study of the Educational Influences of the Typewriter in the Elementary School Classroom. New York, USA, The MacMillan Company.

Yamada, H. (1980). 'A Historical Study of Typewriters and Typing Methods: From the Position of Planning Japanese Parallels.' Journal of Information Processing **2**(4): 175-202.

Zachrisson, B. (1965). Studies in the Legibility of Printed Text. Stockholm, Almqvist & Wiksell.

Zhai, S., Kristensson, P.-O. and Smith, B. A. (2005). 'In Search of Effective Text Input Interfaces for Off the Desktop Computing.' Interacting with Computers **17**: 229-250.

APPENDIX 1 - LAYOUT OF QWERTY UK WINDOWS FULL-SIZED KEYBOARD



APPENDIX 2 - CHILDREN'S PHRASE SET (CPSET)

when we take a bath
there was a crooked man
what did they look like
they all go marching down
I like to eat apples and bananas
morning bells are ringing
jingle all the way
this cold and frosty morning
he was nothing but a pup
give the poor dog a bone
the cupboard was bare
we all scream for ice cream
white clouds on blue sky
thanks for the doughnuts
give them to your sons
down by the bay
she waded in the water
five fat turkeys are we
we spent the night in a tree
over the hills and far away
I caught a fish alive
because he bit my finger
they croaked in the sand
little flowers want to bloom
in the stream so blue
five little speckled frogs
one jumped into the pool
hear the lively song
see what he will say
they went to the church
the owl and the pussycat
looked up to the stars above
would you like to go
put him in a box
jump on your horse
going on a whale watch
underneath your hat
he had ten thousand men
smell as good as new
so she walked right in
riding on a pony
we saw the men and boys
a hole in the ground
there was a sprout
branch on a tree
reaching for the light
I work in a button factory
the mouse ran up the clock
we clean and we scrub
the clock struck one
three white mice
she had so many children
give me a home
when the heavens are bright
light from the glittering stars
I shall miss you
bottom of the deep blue sea
walking down the street
she was fair to see
I dropped my dolly
let your hands go free
I love the rolling hill
on top of a hillside
she plays her guitar
they are so large
like a diamond in the sky
down came the rain
looking out into the night
let your candlelight shine
there are witches in the air
we love you more and more
planted a little watermelon
pulled out the plug
she ruled the others
a skunk sat on a stump
all dressed in black
they touched the sky
the cow jumped over the moon
bright and shiny moon
get your homework done
your friends are my friends
five friends dancing in a line
jumping off the ground
with bright shining faces
through all kinds of weather
he bumped his head
what if the sky should fall
my hand on my head
believe it or not
five little ducks went out to play
the best band in the land
it flew away ever so quickly
right up to your chin
I come from the west

pick them up again
these little hands of mine
the bear went over the mountain
the horse knows the way
it stings the toe
over the ground we go
hear the bell ring
they can make me laugh
first you take the peanut
if all the raindrops
my mouth open wide
put it in the oven
the old man is snoring
he went to bed
will the rain ever stop
I like green frogs
come again another day
little children want to play
gently down the stream
out of my window
would like to sail the ocean
fight with pirates brave and bold
for baby and me
looking in the night
see the barges far ahead
they can hold a crayon
just a boy and a girl
moon shining all around
get out and swim
to see what she could see
we could do a better job
sing along with me
reach my hands way up high
a green and yellow basket
in the month of June
they all rolled over
down by the pond
jumping on the bed
there was an old woman
in the middle of the sea
bumped on the log
this little light of mine
playing with my friends
singing at my school
I look in the mirror
these are my ears
this little piggy stayed at home
give a dog a bone
little ants are marching on
he slept for two hours
we can play on the violin
the leader of the band
all the fine musician
I will make my own shoes
with flags and colours
the horn they blazed away
with his fingers and his thumb
I heard you whispering
at the end of every song
a big grin on his face
till the moon grew dim
the cradle will fall
but no one was there
home came the three bears
so they got married
she broke up the party
knock on the door
tree from a sprout
I had a little turtle
I put him in the bathtub
in came the doctor
he lived in a box
they all decided to race
friends and family began to cheer
the seeds begins to grow
do a dance with me
how I wonder what you are
when the blazing sun is set
this is the zoo
little drop of dew
when the day is light
sailing across the sky
to buy a fat pig
they went so fast
she combed her hair
I went to the field
got to water our horses
they bumped their heads
both began to cry
the big round sun
she added a playroom
the snow has melted
whether we like it or not
watch the leaves tumble
all around the town
the baby on the bus
while you go back to sleep
he can dance alone
letters of the alphabet
with legs like toothpicks
you hold the wand
stepped in the bathtub

marching one by one
wished they all would come back
once I saw an anthill
a little red apple
it looked down at me
to enjoy my lovely dinner
pour the water into the cup
are you sleeping
what fun it is to ride
I caught myself a bumblebee
went to the cupboard
when she got there
teeth in her mouth
the other pointed south
then grew thin again
I married my wife
large ships on blue oceans
hot cross buns
bake me a cake
who stole the cookies
sing me that sweet melody
the part I like best
the cat pleased me
I went down south
down on my knees
he sneezed so hard
where the watermelons grow
back to my home
my mother will say
nothing can compare
she got her ankles wet
when the cook came around
it sure does pay
sleep in the tallest tree
a mighty fine turkey
strut around the barnyard
on a fine spring day
little fishes swimming in the sea
as fast as you can
rocking to the beat
here comes a fish
under the apple tree
they took some honey
plenty of money
sang to a small guitar
she got it from her mother
they sailed away for a year
hand in hand
they danced by the light
went out for a walk
not a tear was in his eyes
I saw a beautiful butterfly
go on a bear hunt
she adored his fur coat
coming to a bridge
we made it home
dance in the tub
all that he could see
other side of the mountain
get the blankets and the food
lying on the cold ground
everyone walks over me
to the tree top
tall as a feather
hold an acorn in your toe
tiny seed planted just right
not a breath of air
I might have a chance
come out tonight
left no room for me
she runs so fast
he eats so much
he rolled around in the mud
always runs the other way
show you a mocking bird
I dare not to go
when the wind blows
sailing off in a wooden shoe
the old moon asked
bringing the fishermen home
a black tassel at the end
make it go away
I wish I had a dinosaur
how often at night
wild flowers in this dear land
she was left all alone
he got bigger and bigger
I have something in my pocket
tired of going everywhere
chased them down an alley
large sack of lamb chops
the prettiest fairy at the ball
I thought I was dreaming
fetched a saucer of milk
deep cave beside the sea
time was running out
the mayor held up a medal
he will give you work
cunning as a fox
we want to play in the park
I used to play in a band
it was glowing and shimmering

the noise came again
a big cloud of dust
shall we go and see
put some water in the kettle
the boat rocked from side to side
like a snowball down a hill
crouch on the ground
little red riding hood
through the looking glass
he bought a new pair of shoes
noise made by a lion
the seed of an oak tree
he always wins at hide and seek
a seat on the beach
peaches and cream for my tea
I was born on a Saturday
she can drive a train
a nice cup of tea
fell in love with a toad
without any warning
at around about noon
under cover of darkness
the lights went off
she was on the floor
reached slowly for a pea
jumping up on my bed
refreshing sea breeze
light the way home
come and have a cuddle
over the hills to the royal castle
go and ask the wizard
she was sulking
the passengers fell silent
looking for a star
at the garden gate
I have everything I want
he would change things
to sing a lullaby
kiss them goodnight
the young foxes ran off
clean up the village
a going home present
they were playing princesses
flowers and fruit trees
sat down on the window seat
silky cushions to lean on
as tall as church tower
your hands stretched up high
he came stomping in
before you catch a cold
hanging by the fire
tell us all about it
hid under the bridge
the flute was very beautiful
made his way to the cliff top
waited for the fisherman
early one morning
like the voice of a bird
rats made from stones
dived back under the sea
she could be our friend
every corner of the cottage
crept up to the bedroom
sound of a baby crying
cats and rats sat on the rocks
from time to time
rocking the buoy to and fro
everyone said so
pushed and shoved and pinched
what are we going to do
sat on the sofa
almost time for dinner
you look tired
he did not slouch
the earth is good to me
you are wonderful boy
a piece of fudge cake
there was no grass left
go to your bedroom
a meadow full of flowers
covered from head to toe
all through the spring
they all leapt up
nowhere else to hide
see you in the morning
on a coach from school
some stopped to look
a lady with long brown hair
you are very clever
stand side by side
they saw a little dog
frog hopped from the pond
they walked by the river
the procession went on
fly round the world
tell everyone to stop
what will become of us
an elephant stampeding through
angry drivers in a traffic jam
they needed a holiday
shout on the plane
there must be something

a lonely seaside cottage
they arrived on the beach
be a good boy
a bag full of things
nearly nodded off to sleep
I knew this would happen
the boat rocked gently
not a single one
king of all the animals
hear what I have to say
the circle was almost complete
he had been eating a carrot
you arrived almost last
there was nothing left
he had a wonderful idea
backwards and forwards
under these nettles
danced round and round
she was quite dizzy
led her down the path
leave you here on your own
there were lots of fairies
I know what to do
fine green grassy field
he skipped and jumped off
middle of the bridge
as thin as can be
strange leathery egg
needed a longer piece of string
it ate all the flowers
I am tired of making tea
it began to rain
we will ride to a safe place
her father and mother
the house was swept away
a fine medal was given
pool made of rubber
I must tell you
feed the ducks at the park
somebody was already sitting there
you poor dear thing
when he had finished
after a little while
jolly little tune
I can cheer him up
the crocodile ran after him
took them all home
laughing with the clown
live secretly in our houses
about the size of a pencil
decide once and for all
it had stopped raining
children were bored
he kept trotting outside
make a cherry cake
she got the vacuum cleaner
made a pot of tea
the wind was whistling
kitchen clock has struck eleven
taking off his raincoat
they heard what had happened
help find that dog
come back after me
that noisy thing
keep away from dangerous things
when the bag is full
on the compost heap
standing in the hall
climbed up on top
whole family was dirty
few muddy paw marks
in a land so far away
beautiful red uniform
he was a cautious man
the witch smiled a smile
a strong young man
the tree is guarded
eyes as big as moons
they will not harm you
as he reached the bottom
quickly filled his pockets
the chest was full
a voice from above
it was a market day
he soon found a room
nice rich prince
watch through the night
every door in town
the fairy godmother was right
in a little bamboo hut
please let me go
scrap of golden paper
their ragged old clothes
went to look outside
cheerful little boys and girls
he ate with his fingers
no more time to play
you have grown too rich
the happiest boy in the world
never said a word

APPENDIX 3 - STUDY PARTICIPANT SUMMARY

																5	6	7	7	7	7	8	9	10
Group Number	Children/Adult	School Name	Date	Participant IDs	No. of Participants	Age range (years)	School year	Num of male	Num of female	Right handed	Left Handed	Demographic Data	CTEQ gathered	Paper-to-screen	Screen-to-Screen	Phrase Set Test	Paper-to-Screen	CE study 1	CE study 2	CE study 3	CE study 4	Cat Meth study	Automisation studies	Final Study
Pilot Studies																								
1	Children	Hesketh Bank	29/06/2006	-	40	7-10	Year 3-5	22	18	-	-	✔		✔		✔						✔		
2	Children	Farrington Moss	14 & 24/02/2007	-	72	5-10	Year 1-5	35	37	-	-	✔		✔			✔					✔		
		Total			**112**			**57**	**55**															
Main Studies																								
3	Children	English Martyres	08/11/2008	C1 - C10	10	6-7	Year 3	6	4	9	1	✔			✔								✔	✔
4	Children	Holme Slack	13/11/2008	C11 - C35	25	9-10	Year 5	16	9	19	6	✔			✔								✔	✔
5	Children	Hesketh Bank	14/11/2008	C36 - C64	29	9-10	Year 5	20	9	26	3	✔			✔								✔	✔
6	Children	Holme Slack	17/11/2008	C65 - C69	5	7-8	Year 3	2	3	4	1	✔			✔								✔	✔
7	Children	English Martyres	20/11/2008	C70 - C94	25	9-10	Year 5	14	11	20	5	✔			✔								✔	✔
8	Children	Manor Road	19/11/2009	C95 - C116	22	9-10	Year 5	14	8	21	1	✔			✔								✔	✔
9	Children	English Martyres	06/07/2010	C117 - C140	24	7-8	Year 3	14	10	18	6	✔			✔				✔				✔	✔
10	Children	Hesketh Bank	06/07/2010	C141 - C163	23	9-10	Year 5	9	14	18	5	✔			✔				✔				✔	✔
11	Children	English Martyres	10/12/2010	C164 - 183	20	8-9	Year 4	12	8	17	3	✔	✔		✔					✔			✔	✔
12	Children	Hesketh Bank	17/03/2011	C184-205	22	8-10	Year 4, 5	10	12	17	5	✔	✔		✔						✔			✔
13	Children	St. Anne's	29 & 30/03/2011	C206-231	26	8-11	Year 4,5,6	17	9	23	3	✔	✔		✔						✔			✔

		Total			**231**			**134**	**97**	**192**	**39**													
Adult Studies																								
14	Adults	Uclan	21/11/2008	S1 - S92	92	18-44	1st UG	80	12	79	13	✔			✔								✔	✔
15	Adults	Uclan	27/09/2010	S93 - S229	137	18-43	1st UG	124	13	114	23	✔	✔		✔			✔					✔	✔
		Total			**229**			**204**	**25**	**193**	**36**													

APPENDIX 4 - WORDS CONTAINED IN OXFORD 3000 LIST

abandon
abandoned
ability
able
about
above
abroad
absence
absent
absolute
absolutely
absorb
abuse
academic
accent
acceptable
accept
access
accident
accidental
accommodation
accompany
account
accurate
accuse
achieve
achievement
acid
acknowledge
acquire
across
act
action
active
activity
actor
actress
actual
actually
ad
adapt
add
addition
additional
address
adequate
adjust
admiration
admire
admit
adopt
adult
advance
advanced
advantage
adventure
advert
advertise
advertisement
advertising
advice
advise
affair
affect
affection
afford
afraid
after
afternoon
afterwards
again
against
age
aged
agency
agent
aggressive
ago
agree
agreement
ahead
aid
aim
air
aircraft
airport
alarm
alarmed
alarming
alcohol
alcoholic
alive
all
allied
allow
ally
almost
alone
along
alongside
aloud
alphabet
alphabetical
already
also
alter
alternative
alternatively
although
altogether
always
amaze
amazed
amazing
ambition
ambulance
among
amount
amuse
amused
amusing
analyse
analysis
ancient
and
anger
angle
angry
animal
ankle
anniversary
announce
annoy
annoyed
annoying
annual
annually
another
answer
anticipate
anxiety
anxious
any
anybody
anyone
anything
anyway
anywhere
apart
apartment
apologise
apparent
apparently
appeal
appearance
appear
apple
application
apply
appoint
appointment
appreciate
approach
appropriate
approval
approve
approving
approximate
approximately
April
area
argue
argument
arise
arm
armed
arms
army
around
arrange
arrangement
arrest
arrival
arrive
arrow
art
article
artificial
artist
artistic
as
ashamed
aside
ask
asleep
aspect
assistance
assistant
assist
associate
associated
association
assume
assure
at
atmosphere
atom
attach
attached
attack
attempt
attempted
attend
attention
attitude
attorney
attract
attraction
attractive
audience
August
aunt
author
authority
automatic
autumn
available
average
avoid
awake
award
aware
away
awful
awfully
awkward
baby
back
background
backward
backwards
bacteria
bad
badly
bag
baggage
bake
balance
ball
ban
bandage
band
bank
bar
bargain
barrier
base
based
basically
basic
basis
bath
bathroom
battery
battle
bay
beach
beak
bear
beard
beat
beautiful
beautifully
beauty
be
because
become
bed
bedroom
beef
beer
before
begin
beginning
behalf
behave
behaviour
behind
belief
believe
bell
belong
below
belt
bend
beneath
benefit
bent
beside
best
bet
better
betting
between
beyond
bicycle
bid
big
bike
bill
billion
bin
biology
bird
birth
birthday
biscuit
bit
bite
bitter

bitterly
black
blade
blame
blank
blind
block
blonde
blood
blow
blue
board
boat
body
boil
bomb
bone
book
boot
border
bore
bored
boring
born
borrow
boss
both
bother
bottle
bottom
bound
bowl
box
boy
boyfriend
brain
branch
brand
brave
bread
break
breakfast
breast
breath
breathe
breathing
breed
brick
bridge
brief
briefly
bright
brilliant
bring
bringing
broad
broadcast
broadly
broken
brother
brown
brush
bubble
budget
build
building
bullet
bunch
burn
burnt
burst
bury
bus
bush
business
businessman
busy
but
butter
button
buy
buyer
by
bye
cabinet
cable
cake
calculate
calculation
call
called
calm
camera
campaign
camp
camping
can
cancel
cancer
candidate
candy
cannot
capable
capacity
cap
capital
captain
capture
car
cardboard
card
care
career
careful
careless
carpet
carrot
carry
case
cash
cast
castle
cat
catch
category
cause
CD
cease
ceiling
celebrate
celebration
cell
cent
centimetre
central
centre
century
ceremony
certain
certainly
certificate
chain
chair
chairman
chairwoman
challenge
challenging
chamber
chance
change
channel
chapter
character
characteristic
charge
charity
chart
chase
chat
cheap
cheaply
cheat
check
cheek
cheerful
cheese
chemical
chemist
chemistry
cheque
chest
chew
chicken
chief
child
chin
chip
chocolate
choice
choose
chop
church
cigarette
cinema
circle
circumstance
citizen
city
civil
claim
clap
class
classic
classroom
clean
clear
clearly
clerk
clever
click
client
climate
climb
chairman
chairwoman
challenge
challenging
chamber
chance
change
channel
chapter
character
characteristic
charge
charity
chart
chase
chat
cheap
cheaply
cheat
check
cheek
cheerful
cheese
chemical
chemist
chemistry
cheque
chest
chew
chicken
chief
child
chin
chip
chocolate
choice
choose
chop
church
cigarette
cinema
circle
circumstance
citizen
city
civil
claim
clap
class
classic
classroom
clean
clear
clearly
clerk
clever
click
client
climate
climb
climbing
clock
close
closed
closet
cloth
clothes
clothing
cloud
club
coach
coal
coast
coat
code
coffee
coin
cold
coldly
collapse
colleague
collect
collection
college
colour
coloured
column
combination
combine
come
comedy
comfortable
comfortably
comfort
command
comment
commercial
commission
commit
commitment
committee
common
commonly
communicate
communication
community
company
compare
comparison
compete
competition
competitive
complain
complaint
complete
completely
complex
complicate
complicated
computer
concentrate
concentration
concept
concern
concerned
concerning
concert
conclude
conclusion
concrete
condition
conduct
conference
confidence
confident
confine
confined

confirm
conflict
confront
confuse
confused
confusing
confusion
congratulate
congratulation
congress
connect
connected
connection
conscious
consequence
conservative
considerable
considerably
consideration
consider
consist
constant
constantly
construct
construction
consult
consumer
contact
contain
container
contemporary
content
contest
context
continent
continue
continuous
contract
contrast
contrasting
contribute
contribution
control
controlled
convenient
conventional
convention
conversation
convert
convince
cook
cooker
cookie
cooking
cool
cope
copy
core
corner
correct
cost
cottage
cotton
cough
coughing
could
council
count
counter
country
countryside
county
couple
courage
course
court
cousin
cover
covered
covering
cow
crack
cracked
craft
crash
crazy
cream
create
creature
credit
crime
criminal
crisis
crisp
criterion
critical
criticism
criticise
crop
cross
crowd
crowded
crown
crucial
cruel
crush
cry
cultural
culture
cupboard
cup
curb
cure
curious
curl
curly
current
currently
curtain
curve
curved
custom
customer
customs
cut
cycle
cycling
dad
daily
damage
damp
dance
dancer
dancing
danger
dangerous
dare
dark
data
date
daughter
day
dead
deaf
deal
dear
death
debate
debt
decade
decay
December
decide
decision
declare
decline
decorate
decoration
decorative
decrease
deep
deeply
defeat
defence
defend
define
definite
definitely
definition
degree
delay
deliberate
deliberately
delicate
delight
delighted
deliver
delivery
demand
demonstrate
dentist
deny
department
departure
depend
deposit
depress
depressed
depressing
depth
derive
describe
description
desert
deserted
deserve
design
desire
desk
desperate
despite
destroy
destruction
detail
detailed
determination
determined
determine
develop
development
device
devoted
devote
diagram
diamond
diary
dictionary
die
diet
difference
different
difficult
difficulty
dig
dinner
direct
direction
directly
director
dirt
dirty
disabled
disadvantage
disagree
disagreement
disappear
disappoint
disappointed
disappointing
disappointment
disapproval
disapprove
disapproving
disaster
disc
discipline
discount
discover
discovery
discuss
discussion
disease
disgust
disgusted
disgusting
dish
dishonest
disk
dislike
dismiss
display
dissolve
distance
distinguish
distribute
distribution
district
disturb
disturbing
divide
division
divorced
divorce
doctor
document
do
dog
dollar
domestic
dominate
done
door
dot
double
doubt
down
downstairs
downward
downwards
dozen
draft
drag
drama
dramatic
draw
drawer
drawing
dream
dress
dressed
drink
drive
driver
driving
drop
drug
drugstore
drum
drunk
dry
due
dull
dump
during
dust
duty
DVD
dying
each
ear
early
earn
earth
ease
easily
east
eastern
easy
eat
economic
economy
edge
edition
editor
educated
educate
education
effect
effective
effectively

efficient
effort
egg
eighteen
eight
eighth
eighty
either
elbow
elderly
elect
election
electrical
electric
electricity
electronic
elegant
element
elevator
eleven
else
elsewhere
email
embarrassed
embarrass
embarrassing
embarrassment
emerge
emergency
emotional
emotion
emphasis
emphasise
empire
employee
employ
employer
employment
empty
enable
encounter
encourage
encouragement
end
ending
enemy
energy
engaged
engage
engine
engineer
engineering
enjoyable
enjoy
enjoyment
enormous
enough
enquiry
ensure
enter
entertain
entertainer
entertaining
entertainment
enthusiasm
enthusiastic
entire
entirely
entitle
entrance
entry
envelope
environmental
environment
equal
equally
equipment
equivalent
error
escape
especially
essay
essential
essentially
establish
estate
estimate
etc
euro
even
evening
event
eventually
ever
everybody
every
everyone
everything
everywhere
evidence
evil
exact
exactly
exaggerated
exaggerate
exam
examination
examine
example
excellent
except
exception
exchange
excited
excite
excitement
exciting
exclude
excluding
excuse
executive
exercise
exhibit
exhibition
existence
exist
exit
expand
expectation
expected
expect
expense
expensive
experienced
experience
experiment
expert
explain
explanation
explode
explore
explosion
export
expose
express
expression
extend
extension
extensive
extent
extra
extraordinary
extreme
extremely
eye
face
facility
fact
factor
factory
fail
failure
faint
fair
fairly
faith
faithful
faithfully
fall
false
fame
familiar
family
famous
fancy
fan
far
farmer
farm
farming
farther
farthest
fashionable
fashion
fasten
fast
fat
father
faucet
fault
favour
favourite
fear
feather
feature
February
federal
feed
fee
feel
feeling
fellow
female
fence
festival
fetch
fever
few
field
fifteen
fifth
fifty
fight
figure
file
fill
film
final
finally
finance
financial
find
fine
finely
finger
finished
finish
fire
firm
firmly
first
fish
fishing
fit
five
fixed
fix
flag
flame
flash
flat
flavour
flesh
flight
float
flood
floor
flour
flower
flow
flu
fly
flying
focus
fold
folding
follow
following
food
football
foot
force
forecast
foreign
forest
forever
for
forget
forgive
fork
formal
former
formerly
form
formula
fortune
forty
forward
foundation
found
four
fourteen
fourth
fox
foxes
frame
freedom
free
freely
freeze
frequent
frequently
fresh
freshly
Friday
fridge
friend
friendly
friendship
frightened
frighten
frightening
from
front
frozen
fruit
fry
fuel
full
fully
function
fundamental
fund
funeral
fun
funny
fur
furniture
further
future
gain
gallon
gamble
gambling
game
gap
garage
garbage
garden
gas
gasoline
gate
gather
gear
general
generally
generate
generation
generous
gentle

gentleman
gently
genuine
geography
get
giant
gift
girlfriend
girl
give
glad
glass
global
glove
glue
goal
god
go
gold
goodbye
good
goods
govern
government
governor
grab
grade
gradual
gradually
grain
gram
grammar
grandchild
granddaughter
grandfather
grand
grandmother
grandparent
grandson
grant
grass
grateful
grave
great
greatly
green
grey
grocery
ground
group
grow
growth
guarantee
guard
guess
guest
guide
guilty
gun
guy
habit
had
hairdresser
hair
half
hall
hammer
hand
handle
hang
happen
happily
happy
hard
hardly
hare
harmful
harm
harmless
hate
hat
hatred
have
headache
head
heal
health
healthy
hear
hearing
heart
heat
heating
heaven
heavily
heavy
heel
he
height
hell
hello
helpful
help
hence
here
her
hero
herself
hers
hesitate
hide
high
highlight
highly
highway
hi
hill
him
himself
hip
hire
his
historical
history
hit
hobby
hold
hole
holiday
hollow
holy
home
homework
honest
honestly
honour
hook
hope
horizontal
horn
horror
horse
hospital
host
hotel
hot
hour
household
house
housing
however
how
huge
human
humorous
humour
hundred
hundredth
hungry
hunt
hunting
hurry
hurt
husband
ice
idea
ideal
identify
identity
if
ignore
I
illegal
ill
illness
illustrate
image
imaginary
imagination
imagine
immediate
immediately
immoral
impact
impatient
implication
imply
importance
important
import
impose
impossible
impressed
impress
impression
impressive
improve
improvement
inability
inch
incident
include
including
income
increase
increasingly
indeed
independence
independent
index
indicate
indication
indirect
individual
indoor
indoors
industrial
industry
inevitable
inevitably
infected
infect
infection
infectious
influence
informal
information
inform
ingredient
in
initial
initially
initiative
injured
injure
injury
ink
inner
innocent
insect
insert
inside
insist
install
instance
instead
institute
institution
instruction
instrument
insulting
insult
insurance
intelligence
intelligent
intended
intend
intention
interested
interesting
interest
interior
internal
international
Internet
interpretation
interpret
interrupt
interruption
interval
interview
into
introduce
introduction
invent
invention
investigate
investigation
invest
investment
invitation
invite
involved
involve
involvement
iron
irritated
irritate
is
island
issue
item
it
itself
its
jacket
jam
January
jealous
jeans
jelly
jewellery
job
join
joint
joke
journalist
journey
joy
judge
judgement
juice
July
jump
June
junior
justice
justified
justify
just
keen
keep
keyboard
key
kick
kid
killing
kill
kilogram
kilometre
kind
kindly
kindness
king
kiss
kitchen
knee
knife
knit
knitted

knitting
knock
knot
know
knowledge
label
lab
laboratory
labour
lacking
lack
lady
lake
lamp
land
landscape
lane
language
large
largely
last
late
later
latest
latter
laugh
launch
law
lawyer
layer
lay
lazy
leader
leading
lead
leaf
league
lean
learn
least
leather
leave
lecture
left
legal
leg
lemon
lend
length
less
lesson
let
letter
level
library
licence
license
lid
lie
life
lift
light
lightly
like
likely
limited
limit
line
link
lip
liquid
listen
list
literature
litre
little
live
lively
living
load
loan
local
located
locate
location
lock
logical
logic
lonely
long
look
loose
loosely
lord
lorry
lose
loss
lost
lot
loud
love
lovely
lover
low
loyal
luck
lucky
luggage
lump
lunch
lung
machine
machinery
mad
magazine
magic
mail
mainly
main
maintain
majority
major
make
male
mall
manage
management
manager
man
manner
manufacture
manufacturer
manufacturing
many
map
March
marketing
market
mark
marriage
married
marry
massive
mass
master
matching
match
mate
material
mathematics
matter
maximum
maybe
May
mayor
meal
meaning
mean
means
meanwhile
measure
measurement
meat
media
medical
medicine
medium
meeting
meet
melt
member
membership
me
memory
mentally
mental
mention
menu
merely
mere
message
mess
metal
method
metre
midday
middle
midnight
might
mild
mile
military
milk
milligram
millimetre
million
millionth
mind
mine
mineral
minimum
minister
ministry
minority
minor
minute
mirror
missing
miss
mistake
mistaken
mixed
mix
mixture
mobile
model
modern
moment
mom
Monday
money
monitor
month
mood
moon
morally
moral
more
moreover
morning
mostly
most
mother
motion
motorbike
motorcycle
motor
mountain
mount
mouse
mouth
movement
move
movie
moving
Mr
Mrs
Ms
much
mud
multiply
mum
murder
muscle
museum
musical
musician
music
must
my
myself
mysterious
mystery
nail
naked
name
narrow
national
nation
naturally
natural
nature
navy
nearby
nearly
near
neat
necessarily
necessary
neck
needle
need
negative
neighbourhood
neighbour
neither
nephew
nerve
nervous
nest
net
network
never
nevertheless
newly
new
news
newspaper
next
nicely
nice
niece
night
nine
nineteen
ninety
ninth
nobody
noise
noisy
none
no
nonsense
normally
normal
nor
northern
north
nose
note
nothing
noticeable
notice
not
novel
November
nowhere
now
nuclear
number
nurse
nut
obey
objective
object
observation
observe
obtain
obviously
obvious

occasionally
occasion
occupied
occupy
occur
ocean
o'clock
October
oddly
odd
offence
offend
offense
offensive
offer
office
officer
officially
official
off
of
often
oh
oil
OK
old
once
one
onion
only
on
onto
opening
openly
open
operate
operation
opinion
opponent
opportunity
opposed
oppose
opposing
opposite
opposition
option
orange
order
ordinary
organisation
organised
organise
organ
originally
original
origin
or
other
otherwise
our
ourselves
ours
outdoor
outdoors
outer
outline
out
output
outside
outstanding
oven
overall
overcome
over
owe
owner
own
pace
package
packaging
packet
pack
page
painful
pain
painter
painting
paint
pair
palace
pale
panel
pan
pants
paper
parallel
parent
park
parliament
particularly
particular
partly
partner
partnership
part
party
passage
passenger
passing
pass
passport
past
path
patience
patient
pattern
pause
payment
pay
peaceful
peace
peak
pencil
penny
pen
pension
people
pepper
perfectly
perfect
performance
performer
perform
perhaps
period
permanent
permission
permit
per
personality
personally
personal
person
persuade
pet
petrol
phase
philosophy
phone
photocopy
photographer
photograph
photography
photo
phrase
physically
physical
physics
piano
pick
picture
piece
pig
pile
pill
pilot
pink
pin
pint
pipe
pitch
pity
place
plain
plane
planet
planning
plan
plant
plastic
plate
platform
player
play
played
pleasant
pleased
please
pleasing
pleasure
plenty
plot
plug
plus
pocket
poem
poetry
pointed
point
poisonous
poison
pole
police
policy
polish
polite
politically
political
politician
politics
pollution
pool
poor
pop
popular
population
pore
port
pose
position
positive
possession
possess
possibility
possible
possibly
post
potato
potential
pot
pound
pour
powder
powerful
power
practically
practical
practice
practise
praise
prayer
pray
precisely
precise
predict
preference
prefer
pregnant
premises
preparation
prepared
prepare
presence
presentation
present
preserve
president
press
pressure
presumably
pretend
pretty
prevent
previous
price
pride
priest
primarily
primary
prime
prince
princess
principle
printer
printing
print
priority
prior
prisoner
prison
private
prize
probable
probably
problem
procedure
proceed
process
produce
producer
production
product
professional
profession
professor
profit
programme
program
progress
project
promise
promote
promotion
promptly
prompt
pronounce
pronunciation
proof
properly
proper
property
proportion
proposal
propose
prospect
protection
protect
protest
proudly
proud
prove
provided
provide
providing
publication
publicity
public
publishing
publish
pub
pull
punch
punishment
punish
pupil
purchase
purely
pure
purple
purpose

pursue
push
put
qualification
qualified
qualify
quality
quantity
quarter
queen
question
quickly
quick
quiet
quite
quit
quote
race
racing
radio
rail
railroad
railway
rain
raise
range
rank
rapid
rarely
rare
rate
rather
raw
reach
reaction
react
reader
reading
read
ready
realistic
reality
realise
really
real
rear
reasonable
reasonably
reason
recall
receipt
receive
recently
recent
reception
reckon
recognition
recognise
recommend
recording
record
recover
red
reduce
reduction
reference
refer
reflect
reform
refrigerator
refusal
refuse
regarding
regard
regional
region
register
regret
regularly
regular
regulation
reject
related
relate
relation
relationship
relatively
relative
relaxed
relaxing
relax
release
relevant
relief
religion
religious
rely
remaining
remain
remains
remarkable
remark
remember
remind
remote
removal
remove
rented
rent
repair
repeated
repeat
replace
reply
report
representative
represent
reproduce
reputation
request
requirement
require
rescue
research
reservation
reserve
resident
resistance
resist
resolve
resort
resource
respect
respond
response
responsibility
responsible
restaurant
restore
rest
restricted
restriction
restrict
result
retain
retired
retirement
retire
return
reveal
reverse
review
revise
revision
revolution
reward
rhythm
rice
rich
ride
rider
ridiculous
riding
rid
rightly
right
ring
rise
risk
rival
river
road
rob
rock
role
roll
romantic
roof
room
root
rope
roughly
rough
rounded
round
route
routine
row
royal
rubber
rubbish
rub
rudely
rude
ruined
ruin
ruler
rule
rumour
runner
running
run
rural
rush
sack
sadly
sadness
sad
safely
safe
safety
sailing
sailor
sail
salad
salary
sale
salt
salty
same
sample
sand
satisfaction
satisfied
satisfying
satisfy
Saturday
sauce
save
saving
say
scale
scared
scare
scene
schedule
scheme
school
science
scientific
scientist
scissors
score
scratch
scream
screen
screw
seal
search
sea
season
seat
secondary
second
secretary
secret
section
sector
secure
security
seed
seek
seem
see
selection
select
self
sell
senate
senator
send
senior
sense
sensible
sensitive
sentence
separated
separately
separate
separation
September
series
seriously
serious
servant
serve
service
session
set
settle
seven
seventeen
seventy
several
severe
sewing
sew
sex
sexual
shade
shadow
shake
shallow
shall
shame
shaped
shape
share
sharply
sharp
shave
sheep
sheet
shelf
shell
shelter
she
shift
shine
shiny
ship
shirt
shocking
shock
shoe
shooting
shoot
shopping
shop
shortly
short
shot
shoulder
should
shout
shower
show
shut
shy
sick

side
sideways
sight
signal
signature
significantly
significant
sign
silence
silent
silk
silly
silver
similarly
similar
simple
simply
sincerely
sincere
since
singer
singing
single
sing
sink
sir
sister
site
sit
situation
six
sixteen
sixty
size
skilful
skilled
skill
skin
skirt
sky
sleep
sleeve
slice
slide
slightly
slight
slip
slope
slowly
slow
small
smart
smash
smell
smile
smoke
smoking
smoothly
smooth
snake
snow
soap
social
society
sock
softly
soft
software
soil
soldier
solid
solution
solve
somebody
somehow
someone
some
something
sometimes
somewhat
somewhere
song
son
soon
sore
sorry
sort
so
soul
sound
soup
source
sour
southern
south
space
spare
speaker
speak
specialist
specially
special
specifically
specific
speech
speed
spelling
spell
spend
spice
spicy
spider
spin
spirit
spiritual
spite
split
spoil
spoken
spoon
sport
spot
spray
spread
spring
square
squeeze
stable
staff
stage
stair
stamp
standard
stand
stare
star
start
statement
state
station
statue
status
stay
steady
steal
steam
steel
steep
steer
step
stick
sticky
stiff
still
sting
stir
stock
stomach
stone
stop
store
storm
story
stove
straight
strain
strangely
stranger
strange
strategy
stream
street
strength
stressed
stress
stretch
strictly
strict
strike
striking
string
striped
stripe
strip
stroke
strong
structure
struggle
student
studio
study
stuff
stupid
style
subject
substance
substantially
substantial
substitute
succeed
successful
success
such
suck
suddenly
sudden
suffering
suffer
sufficient
sugar
suggestion
suggest
suitable
suitcase
suited
suit
sulk
sulked
sulking
summary
summer
sum
Sunday
sun
superior
supermarket
supply
supporter
support
suppose
surely
sure
surface
surname
surprised
surprise
surprising
surroundings
surrounding
surround
survey
survive
suspect
suspicion
suspicious
swallow
swearing
swear
sweater
sweat
sweep
sweet
swelling
swell
swimming
swim
swing
switch
swollen
symbol
sympathetic
sympathy
system
table
tablet
tackle
tail
take
talk
tall
tank
tape
tap
target
task
taste
taxi
tax
teacher
teaching
teach
team
tear
tea
technical
technique
technology
telephone
television
tell
temperature
temporary
tendency
tend
tension
ten
tenth
tent
term
terrible
terribly
test
text
thanks
thank
than
that
the
theatre
theirs
their
theme
themselves
them
then
theory
therefore
there
they
thickly
thickness
thick
thief
thing
thinking
think
thin
third
thirsty
thirteen
thirty
this
thoroughly
thorough
though
thought
thousand
thousandth
thread
threatening

threaten
threat
three
throat
throughout
through
throw
thumb
Thursday
thus
ticket
tidy
tie
tightly
tight
timetable
time
tin
tiny
tip
tired
tire
tiring
title
today
toe
together
toilet
tomato
tomorrow
tone
tongue
tonne
ton
tool
tooth
too
topic
top
totally
total
to
touch
tough
tourist
tour
towards
tower
town
toy
toys
trace
track
trade
trading
traditional
tradition
traffic
training
train
transfer
transform
translate
translation
transparent
transportation
transport
trap
traveller
travel
treatment
treat
tree
trend
trial
triangle
trick
trillion
trip
tropical
trouble
trousers
truck
true
truly
trust
truth
try
tube
Tuesday
tune
tunnel
turn
TV
twelve
twenty
twice
twin
twisted
twist
two
type
typically
typical
tyre
ugly
ultimately
ultimate
umbrella
unable
unacceptable
uncertain
uncle
uncomfortable
unconscious
uncontrolled
underground
underneath
understanding
understand
under
underwater
underwear
undo
unemployed
unemployment
unexpected
unfair
unfortunately
unfortunate
unfriendly
unhappy
uniform
unimportant
union
unique
united
unite
unit
universe
university
unkind
unknown
unless
unlikely
unlike
unload
unlucky
unnecessary
unpleasant
unreasonable
unsteady
unsuccessful
untidy
until
unusually
unusual
unwilling
upon
upper
upsetting
upset
upstairs
up
upwards
upward
urban
urgent
urge
used
useful
useless
user
use
usually
usual
us
vacation
valid
valley
valuable
value
van
variation
varied
variety
various
vary
vast
vegetable
vehicle
venture
version
vertical
very
via
victim
victory
video
view
village
violence
violently
violent
virtually
virus
visible
vision
visitor
visit
vital
vocabulary
voice
volume
vote
wage
waist
waiter
wait
wake
walking
walk
wallet
wall
wander
want
warmth
warm
warning
warn
war
was
washing
wash
waste
watch
water
wave
way
weakness
weak
wealth
weapon
wear
weather
website
web
wedding
Wednesday
weekend
week
weight
weigh
welcome
well
western
west
wet
we
whatever
what
wheel
whenever
when
whereas
wherever
where
whether
which
while
whisper
whistle
white
whoever
whole
whom
whose
who
why
widely
wide
width
wife
wildly
wild
willing
will
window
wind
wine
wing
winner
winning
winter
win
wire
wise
wish
withdraw
within
without
with
witness
woman
wonderful
wonder
wooden
wood
wool
word
worker
working
work
world
worried
worrying
worry
worse
worship
worst
worth
would
wounded
wound
wrapping
wrap
wrist
writer
write
writing
written
wrongly
wrong
yard
yawn
yeah
year
yellow
yesterday
yes

yet
young
yourself
yours
your
youth
you
zero
zone