ExpECT: An Expanded Error Categorisation Method for Text Input

Akiyo Kano
Department of Computing
University of Central
Lancashire, Preston,
PR1 2HE, UK
+44 (0)1772 89 5300

akano@uclan.ac.uk

Janet C. Read
Department of Computing
University of Central
Lancashire, Preston,
PR1 2HE, UK
+44 (0)1772 89 3285

Jcread@uclan.ac.uk

Alan Dix
Computing Department
Lancaster University
Lancaster
LA1 4YR, UK
+44 (0)1524 510319

alan@hcibook.com

I. Scott MacKenzie
Department of Computer
Science and Engineering
York University, Toronto
Ontario, Canada, M3J 1P3
+1 416 736 2100

mack@cse.yorku.ca

ABSTRACT

This paper describes an empirical study on typing errors made by children during a text copy exercise. The literature on text input errors is first examined, focussing on studies of errors that occur during keyboard typing. A study of errors made by children during typing is described and the results from this study are analysed using visual inspection and already published error categorisation methods. These methods are compared with respect to the types and number of errors categorised and uncategorised. We identify and define new kinds of typing errors and use these, together with previously defined error types, to outline an expanded and more detailed method (ExpECT) for the classification of typing errors. ExpECT is compared with the previously examined categorisation methods and is shown to be a more thorough and broader method for the analysis of typing errors.

Categories and Subject Descriptors

H.5.5 **[Information Interfaces and Presentation (e.g., HCI)]**: User Interfaces - *input devices and strategies, interaction styles, evaluation/methodology.*

General Terms

Experimentation, Human Factors.

Keywords

Text input, keyboards, error analysis, typing error, error categorization, children.

1. INTRODUCTION

This is work from a larger project investigating causal relationships between visual processing skills, language processing skills, and typing behaviours for children. It is hypothesised that these relationships offer a low cost, low effort indicator for the early identification of difficulties in language processing and visual processing for children. A second aim is to gather and code data demonstrating how children carry out typing tasks.

In keyboard-based text input evaluations, we use samples of typing to examine error categorisation. Most studies on text input are comparative, whereby one independent variable, such as input method, is manipulated whilst participants enter text. Comparisons are undertaken in character or word error rates, speed of entry, user preference, and so on. Our study takes a different approach by examining how participants differ in entering phrases under the same condition. This methodology follows from the different purpose of our research. In many studies the aim is to understand the difference between several techniques, or often to 'prove' that a new technique is superior. In contrast our ultimate aim is to investigate whether particular typing behaviours relate to visual processing and language processing difficulties. In order to identify the aspects of typing mistakes that relate to these processing difficulties, the full range of typing errors that can occur during typing must be captured and categorised for analysis.

Children were used as participants in the empirical study described here for two reasons. Firstly, children are the main focus of the larger project, of which this study is a part. Secondly, children are 'more similar'; there is a much more variability in how adults type. Some adults can barely type, while others may be fast hunt-and-peck typists or even touch typists. Young children are less likely to have such wide variations in terms of typing experience. Although derived from a study using children, the findings in this paper and the new categorisation methods are applicable to both adults and children.

2. BACKGROUND

Keyboard based text input has been studied in HCI for several decades; however, given that text input is a very common activity for the average user, it is somewhat surprising that there are relatively few significant studies. When word processing was first introduced, there were several studies on typing [2, 6, 16]. As the GUI interaction became popular, the studies on text input declined. More recently, there is an upsurge of interest due to the introduction of different text input methods (particularly in mobile computing), but studies that just focus on the keyboard are rare.

Even rarer are studies on children entering text at keyboards. Three papers by the authors address the design of text copy tasks for children [5], comparing four different input methods with young children [15], and the keyboard behaviours of teenagers [14]. These aside, most studies focus on the use of software, such as word processors, and lack attention to detail on the typing process.

2.1 Text Input Demystified

Text input is the general process by which text is inputted to a computer. There are several methods, which are well documented in [12]. One method is a keyboard-based text input, which uses key presses to input text. The keyboard can be full sized (like a QWERTY keyboard) or can be reduced (as found on mobile phones). Full-sized keyboard text entry is generally referred to as typing, a process that has been well studied. Cooper [1] identifies five stages for typing: character recognition, storage, motor activity, keystroke, and feedback. All must occur without error for the process to be successful.

2.2 Evaluation of Text Input

There are two approaches to text input evaluation. The first is based on error rates and participant efficiency. In this model, counts are made of the number of errors committed, the number of errors corrected and uncorrected and the effort expended by the participant, measured as keystrokes, or keystrokes per character and as time taken. The second approach is to consider the different errors made and to classify them in some way.

Central to both approaches is the comparison of two text strings, the intended text (that which was being copied or was intended to be created) (IT) and the transcribed text (that which appears) (TT). These two strings are lined up using an MSD (Minimum String Distance) algorithm [17]. Once the strings are aligned, the errors can be counted, and later investigated. A base error rate is derived by counting the minimum number of primitives (insertions, deletions or substitutions) needed to transform one string into the other. This error rate thus shows how far the transcribed text is from the intended text.

Another metric represents effort or efficiency [7]. KSPC (Keystrokes per Character) is the number of key presses, on average, to generate each character in a text phrase. Where capitalisation is not included, for perfect (error free) performance, the KSPC for a QWERTY keyboard is 1; for QWERTY, a value greater than 1 implies erroneous behaviour (in reduced keyboards, like those found on mobile phones, the 'optimal' KSPC is greater than 1). In QWERTY typing, a large KSPC indicates that a participant either made, for example, few errors that were costly to fix or many errors that were easy to fix; it does not distinguish between these two scenarios.

In the same paper [7], several other metrics were introduced to better understand errors. Total Error Rate (TER) is the percentage of errors, irrespective of whether an error was corrected or not. Not Corrected Error Rate (NCER) is the percentage of errors left unfixed and Corrected Error Rate (CER) is the percentage of errors fixed. These three metrics indicate how many errors were made overall, and how many were left fixed or unfixed, but do not shed any light as to why the errors were made. In addition, the metrics treat all errors the same and simply provide a summation that allows for comparison across instances.

3. CATEGORISATION OF ERRORS

As noted, in addition to raw data about the number of errors made, researchers have sought to determine what types of errors are made, in an attempt to understand why they were made with the hope that they might be prevented. The categorisation of text input errors has been a long running theme in HCI that began in the years before the computer. These studies are summarised here in a chronological order moving from early work on typing to more recent studies with PCs.

3.1 Typewriter Studies

In the early 20th Century, when the typewriter 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 and desiring to improve keyboards and the teaching of keyboarding skills, several individuals carried out studies involving large samples of transcriptions. Lessenberry [6] took 60,000 erroneously typed letters and created letter confusion matrices according to the intended letter and what was actually struck. This analysis provided information that showed how likely a letter would be struck for another letter and found that vowels were often confused with other vowels. Lesssenberry, in reporting these results, admitted that the contextual information are 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. Smith [16] also demonstrated that such letter-error counts are artificial and on their own, provide little value, as the majority of typing errors are sequence errors that are unrelated to isolated key presses.

In one of the earliest review papers on typing errors, Dvorak et al. [2] summarised error categorisation work by their contemporaries and, as a result of this work, defined several error types, which consider not only errors of isolated letters but errors involving several letters or even entire words. Most notably, Dvorak et al. cite White [19] who analysed 20,623 typing errors on QWERTY keyboards and categorised them into 10 error types that concerned both the letter level and word level; these are shown in Table 1.

Table 1. Errors found in White [19].

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%
Repeating Words	1%
Omitting Words	1%

In their own classification, Dvorak et al. [2] define several other error types. These include Adjacent Errors, which are errors where the keystroke is falsely directed to an adjacent key, such as typing 'amvition' for 'ambition'. These adjacent errors correspond to 60% of Lessenberry's errors. A second error type is Copy-Reading Errors, which occur when the word was (assumed to be) read incorrectly from the presented text. An example given by Dvorak et al. is the typing of 'admiration' for 'ambition'. Two transposition errors are described, the first is the basic Transposition error which is defined as an interchange of two strokes; a Transposed Doubling error is one where two such errors occur, for instance in the typing of 'thses' for 'these'. Although lacking formal definitions in the book, other classifications including Omitting Word, Substitution, and Omission, are also mentioned.

3.2 Word Processing Studies

The word processor offered new opportunities for the study of text input, as the text so created was easier to study. In 1983, a major work in studying typing was carried out by Cooper et al., [1] in which Gentner et al. [3], formally defined and listed nine different error types as summarised in Table 2.

Table 2. Summary of typing errors defined in [3].

Error	Definition
Misstrokes	An error which can be traced to inaccurate motion of the finder, 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.
Interchange across I letters	Two non-consecutive letters are switched with I letters intervening (I>0).
Migration across M letters	One letter moves to a new position, with M letters intervening between its correct position and its end position (M>1).
Omission	When a letter in a word is left out.
Insertion	An extra letter is inserted into a text. Some insertions can be classified as misstrokes.
Substitution	When the wrong letter is typed in place of the correct letter.
Doubling Error	Word containing a repeated letter and the wrong letter is doubled instead.
Alternating Error	When a letter alternates with another but the wrong alternation sequence is produced.

Note that the term Transposed Doubling Error defined by Dvorak et al. [2] was described by Gentner et al. as an 'Alternating Error', which perhaps better describes the nature of the error.

In the same Cooper et al.. book, Grudin [4] carried out a transcription study involving expert typists and 70 beginner typists at high school and categorised their typing errors into Substitution, Insertion, Omission, Transposition and Other. He studied in particular two types of substitutions. First were substitution errors amongst homologous letters (letter pairs which are pressed by the same finger on the same position but on different hands). Second were substitutions of adjacent letters on the keyboard. He concluded that both were likely to be caused by error in the control of the typing fingers. He also used video recordings to understand whether typists intended to press the erroneous key or accidentally hit it.

3.3 Errors and Intention

In investigating video recordings, Grudin [4] was attempting to discover something of the intention of the typists. Intention is a difficult for all studies of human behaviour and for text input, the tale is no different. Norman [13], was one of the first to consider intention in respect of errors, describing a 'mistake' as an error in the intention of the user, and a 'slip' as an error where the intention was correct but an error was made in carrying out the intention. In the same paper, he further describes slips based on the Activation-Trigger-Schema System (ATS), which assumes that, to perform a task, several schemas

are activated and are selected by a triggering mechanism. This triggering mechanism requires appropriate conditions to be satisfied for the schema to be operational. He categorised slips depending on where in this model the error occurred.

In a more recent study, Read et al. [15] extended the idea of errors as mistakes and slips in an application of typing error categorisation to children's typing. Read et al. 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. The error types were categorised based on slips, mistakes and also hardware related errors. These errors, outlined in Table 3, include errors that would not occur in QWERTY typing activities.

Table 3. Categorising errors defined in Read et al. [15] as slips, mistakes [13] and hardware related errors.

Error Types	Example	Cause
Cognition Error	Child misreads a word or cannot distinguish letters.	Mistake
Spelling Error	Child misspells words or mispronounces a word that they know.	Mistake/ Slip
Selection Error	Child picks 'l' for 'i'.	Hardware
Construction Error	Child cannot form the letter or word correctly. In handwriting, 'a' may look like'd'. In speech, 'dragon' becomes 'dwagon'.	Slip
Execution Error	The child presses for too long, fails to click or hits the adjacent character. The microphone or pen may be out of range.	Hardware
Software Induced Error	The software misrecognises the word or character.	Hardware

In a study focussed on typing, Read and Horton [14] carried out a text copying task with eighteen 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' instead of 'over the', and Unknown Errors (U) are those errors where there is no obvious reason. NT errors were 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 that 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'.

4. METHODS AND TOOLS FOR COUNTING ERRORS

Ambiguity and accuracy in classification are two problems facing researchers in this field. Ambiguity can never be eliminated but algorithms, based on probabilistic processes and contextual information, can reduce the uncertainty about what caused an error. In a typical text input study, large amount of data is investigated and so, to reduce error, several tools and methods have been devised to count, classify and report errors. Much of the work in this area has been lead by MacKenzie and Soukoreff [7, 8, 10, 18] who developed several automated tools to correctly identify the TER, CER and NCER [18], Their analyser takes the Presented phrase (PT), the Input Stream (IS) and the final Transcribed text (TT) and categorises each letter as Correct (C), Incorrect and Fixed (IF) or Incorrect and Not Fixed (INF). However their tool does not categorise the errors into error types. The analyser is embedded in a software product that shows single phrases for participants to input via the specified input method. Once a phrase is entered, a second phrase is shown and the participants continue until all the phrases that they have been allocated have been completed.

Wobbrock and Myers [20] took MacKenzie and Soukoreff's work further and defined a method in which IS is compared with PT and categorises each letter in the IS as either Substitution, Omission, Insertion or No-Error. Their analyser also determines whether the letters were 'corrected' or not, for example, a substituted letter becomes either a Corrected Substituted letter or an Uncorrected Substituted letter depending on whether an attempt was made to correct it. The authors noted that the term 'corrected' refers to the fact that the character was backspaced, likely in an attempt to correct a mistake, rather than the erroneous letter actually being fixed.

In an effort to reduce the amount of ambiguities in categorising an error Wobbrock and Myers [20] introduced four well-reasoned assumptions about participant behaviour. Participants proceed sequentially through PT as they enter IS, they only insert or omit only one character at a time, they backspace accurately, and intentionally and that those letters omitted in TT are also omitted in IS. The limitation of this method is that it only accepts a backspace as the editing method, even if a mouse or the directional keys are available for the participant. To reflect this, they placed a restriction on their participants that limited their phrase copying task program to only allow for backspaces and disabled the directional keys. Their analyser also does not allow for functional key presses such as ctrl and shift and, as with the Soukoreff and MacKenzie [18] program, the participants are only shown one line of phrase at a time.

4.1 Limitations of Existing Methods

The error categorisation methods and tools discussed in this paper have varying limitations.

The studies of Lessenberry and Gentner only carried out letter-by-letter comparison and thus did not count any word errors such as omitted words. The early works of Lessenberry, Dvorak and Grudin were also restricted by technology and could only consider the finally transcribed text.

Due to how phrases are presented to participants, the categorisation methods of Wobbrock and MacKenzie do not account for errors made across the phrases (e.g., omitted phrases). Other errors typically not caught include when the return key is hit mid-sentence, characters backspaced in the process of fixing an error or by accident, deletions caused by pressing the backspace key too many times. It is difficult to exhaustively list all the error types that these methods do not take into account, since little work has been carried out on capturing the full range of errors that are made during a copy-typing task, particularly those carried out by children.

5. THE STUDY

A three-day study involving 112 children from two local junior schools was carried out to collect a large amount of phrase copying task data for analysis. Phrase copying was preferred as it allows us to know what the participant intended to type [5, 11, 20].

5.1 Participants and Apparatus

There were 57 boys and 55 girls, aged between 5 and 10 years. 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 displays) as shown in Figure 1.

Figure 1. Set up of the experiment.

5.2 Design

Rather than using an experimental text input interface (as used by Wobbrock and MacKenzie) in this study, children copied phrases shown on paper into Notepad™ via a standard QWERTY keyboard. Based on findings of previous studies [5], younger children were only asked to type 5 phrases each, while the older children typed 10. 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 (Text Entry Phrase Set [9]) and CPSet (Children's Phrase Set [5]). The 100 resulting phrases were each presented approximately 10.6 times. The phrase shown first was randomized to eliminate learning effects. In all, 1060 phrases were shown to the children.

5.3 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 their 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®, which 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.

The University of Central Lancashire's Faculty of Science and Technology Ethics Committee approved this study. Parental consent was obtained via each school prior to the study. The children were also told at the start, and reminded throughout the study, that they were free to stop participating if they wished to do so at any time.

5.4 Analysis Method

5.4.1 Manual Classification

Firstly, a manual analysis of the data was carried out to gauge the total number of errors made by 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 presented text (PT) was compared with the complete Input Stream (IS). Any errors found were noted for whether an attempt at fixing the error was made, as shown below:

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

In this example, where a backspace is indicated by '<', we see 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. We also see an uncorrected error where 'my' was typed without the 'y'. The word 'home' is typed as 'house' and this is considered, in this classification, as one corrected word error.

5.4.2 Gentner et al. Classification

A second visual classification was then carried out based on the eight error types defined by Gentner et al. [3]. This used only the transcribed text (TT) ignoring the Input Stream. The classifications were Transposition, Interchange, Migration, Omission, Insertion, Substitution, Doubling Error and Alternating Error. For example, in the previous example:

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

There is only one error remaining in TT: the omission of 'y' in 'my'.

5.4.3 Wobbrock and Myers Classification

The third method carried out classified errors as defined in Wobbrock and Myers [20]; Corrected Omission, Uncorrected Omission, Corrected Insertion, Uncorrected Insertion, Corrected Substitution, Uncorrected Substitutions, and Corrected No-Error.

As the children had not used the Wobbrock & Myers [20] StreamAnalyser, the researcher had to input the Input Streams along with the Presented Text. When there are several possible combinations of error types, the program offers all possible combinations and the researcher chose the least costly combination.

5.5 Results

5.5.1 Manual Inspection of Key Logs

Out of the 1060 phrases shown, the children omitted 30 phrases, attempted to copy 1030 phrases (25531 letters), of which they left seven incomplete. The input streams of text, from the key logs, were manually analysed and 2312 errors were found (Table 4). 125 were word-level errors and the remaining 2187 were letter-level errors. Of the 2312 errors, 99% were fixable errors (i.e. pressing a function error was not fixable). However, only half of the errors (49.4%) were fixed. When a child made an error, 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.

Table 4. 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

5.5.2 Gentner et al. Classification

This analysis identified 1327 errors in the transcribed text as summarised in Table 5. 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 5. Summary of Gentner et al. [3] 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	

5.5.3 Wobbrock and Myers Classification

This analysis found 2490 errors in total. A summary of the counts is shown in Table 6.

Table 6. Summary of error counts found by Wobbrock and Myers [20] classification.

	Corrected	Uncorrected	Total
Insertion	403	555	958 (38.5%)
Omission	238	518	756 (30.4%)
Substitution	433	343	776(31.13%)
Total	1074 (43.13)	1416 (56.87%)	2490

There was a more equal 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 L/R 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 that were 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.

5.6 Comparison of the Categorisation Methods

The categorisation methods defined by Gentner et al. [3] and Wobbrock and Myers [20] were compared with the manual inspection of the key logs, to find out to what extent each method ruled out certain error types, and how the different definitions of errors affected the result. Table 7 summarises the number of errors found by each method.

Table 7. 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 (51.9%)	203 (8.8%)

5.6.1 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 the transcribed text. In addition, due to the letter-only based nature of the definitions of error types in [3], 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.

5.6.2 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 & Myers' accompanying program TextTest, they would not be able enter any text once the Return key was hit. 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 letter-based nature of the error type definitions, one word-level error in the manual classification was considered as multiple single-letter errors.

The classification does not differentiate between a correctly typed character deleted during editing from one deleted by

error. If however we were to make this distinction between the two types of error, 114 out of the 391 Corrected No-Error occasions were erroneous. Wobbrock and Myers [20] 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 3 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 a phrase-copying task using a QWERTY keyboard. Although Gentner et al. [3] 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' [20] categorise more errors by considering those that were corrected in the Input Stream. However, it loses the detail of what errors were made due to featuring a reduced number of error types.

5.7 Discussion

Although both Gentner et al. [3] and Wobbrock and Myers' [20] 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 reach the final aim of this project – understanding how and why children make typing mistakes – we needed to devise a method that categorises as many typing errors as possible, and in as fine detail as possible.

6. EXPECT - A NEW ERROR CLASSIFICATION METHOD

We have devised a new classification method to categorise typing errors from copy-typing tasks as thoroughly as possible. It also categorises errors into broader ranges based on the error patterns, providing a detailed analysis on how the errors are made.

In this new method, we have merged error types in the literature with several new errors types not previously defined. Most significantly, we have extended on White [19] by introducing more word level errors and even phrase level errors. These error types are concerned with errors on a larger scale.

Next To and Close Errors [14] 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 [14] have 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 have been subcategorised into two types according to the source of the word that is substituting or inserted. This is to reflect the fact that participants and, especially in our case, 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.

Several error types that seem unlikely in experienced adult typists but are peculiar to children have been defined. These

include adding a full stop after every phrase (ISy), holding down the keys for too long [15] (ExE), and capitalising every initial letter in a phrase (CE).

Finally, we have differentiated between Corrected No-Errors that occur erroneously, to those that are erased intentionally during an edit of the text. Only those that were done erroneously are considered as an error in this method.

6.1 Error Type Defined

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 8.

Table 8. Summary of new error types.

	Omission	Substitution	Insertion	Other
Letter	OL OS	AE CE CT-S DE IE NT-S SL TE ME	CT-Mu DL DS ExE 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	

6.2 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.

6.2.1 Omission Error Types

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)
```

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 MS errors are shown below:

```
thanksfor (thanks for)
doorsare (doors are)
```

6.2.2 Substitution Error Types

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)
```

Transposition Error (TE)

The definition of Transposition errors (TE) remains unaltered from Gentner et al.. [3]: '*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)
```

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 2 shows that, if the intended key is 'G', then the keys 'F' and 'H' are classified as NT keys.
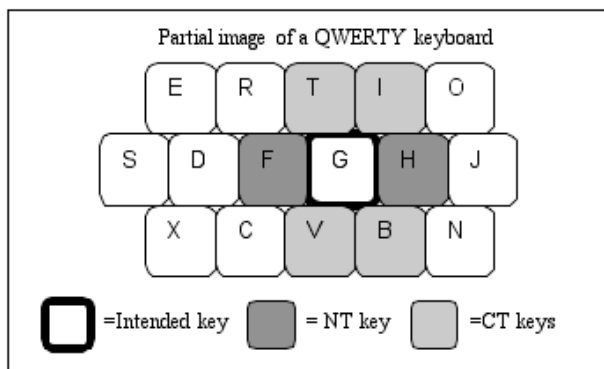

Figure 2. NT and CT keys for 'G' on a QWERTY keyboard.

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.

Close To error – Substitution (CT-S)

'Close To' keys to an intended key are those keys that are 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 2 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.

Capitalisation Error (CE)

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.

Alternating Error (AE)

The definition for Alternating Error (AE) remains unchanged from Gentner et al. [3]: '*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.

Doubling Error (DE)
The definition for a Doubling Error (DE) remains unaltered from Gentner et al. [3]: *'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.

Interchange (IE)
The definition of an Interchange Error (IE) across I letters remains unaltered from Gentner et al. [3]: *'Two non-consecutive letters are switched with I letters intervening (I>0)'*

Migration (ME)
The definition of a Migration Error (ME) across M letters remains unaltered from Gentner et al. [3]: *'One letter moves to a new position, with M letters intervening between its correct position and its end position (M>1)'*.

```
orrecection (correction)
```

6.2.3 Insertion Error Types
Inserted Letter (IL)
When an extra letter (not a duplicate of the pervious letter) is inserted, it is classified as an IL (Inserted Letter) error. Some examples include:

```
hern (her)
docktor (doctor)
```

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.

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 classified as an NT-Mu error.

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.

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)
```

Duplicated Space (DS)
An error is classified as an DS (Duplicated Space) error if two spaces are typed when only one space is shown in the presented text.

```
all  that  he  could (all that he could)
these  cookies (these cookies)
```

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.

Inserted Function (IF)
If a function key, such as Control and 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.

Corrected No-Error by Error (CNE(error))
Wobbrock and Myers [20] 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).

6.3 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.

6.3.1 Omission Error Types
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  (they all go marching)
one zero blast off (two one zero blast off)
```

6.3.2 Substitution Error Types
Substituted Word – word from Another place on the phrase sheet (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)
```

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)
```

6.3.3 Insertion Error Types
Inserted Word – word from Another place on the phrase sheet (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)
```

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.

Dupilicated 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)
```

6.4 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.

6.4.1 Omission Error Types
Omitted Phrase (OP)
When an entire phrase is omitted, it is classified as an Omitted Phrase (OP).

6.4.2 Substitution Error Types
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.

6.4.3 Insertion Error Types
Duplicated Phrase (DP)
Duplicated Phrase error (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.

6.5 Other Error Types
Enter Error (EE)
When the Return key is pressed in a place other than at the end of a phrase, it is classified as an EE (Enter Error). Some examples of EE are:

```
with bright shining [ENTER]
```

```
faces
(with bright shining faces)
```

Execution Error (ExE)
Defined by Read et al. [15], 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)
```

Unknown (U)
When the error does not fit any category 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 (???)
```

7. 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 5.4.1. Table 9 shows a summary.

Table 9. Summary of errors categorised using ExpECT.

Error Type	Frequency	Error Type	Frequency
Ambiguous	19	IS	89
AE	0	ISy	170
CE	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	4
IL	107	Total	2312

With the new classification there were just 19 ambiguities. As the Input Stream was used as the source of the errors (as opposed to the final transcribed text), many errors were noticed and fixed by the participants half way 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.

In addition, the method does not yet differentiate between errors which are slips and those that are mistakes as defined by Norman [13]. For example, knowing if the error was caused because the participant really thought 'doctor' was spelt with an 'e' would allow further understanding to why the error was caused. However, to understand this, additional tests with each participant would be required, for instance a written spelling test to establish whether the participant spelt 'doctor' as 'docter' away from the computer.

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 automatically, 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. CONCLUSION

This study has reviewed the typing errors as defined in the literature and applied two contrasting methods of categorising typing errors created by children during a copy-typing task. We discovered that the methods and definitions lacked in some aspects of classifying all errors. 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. We have therefore introduced a new typing error classification method 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. This will undoubtedly require some additional work, for example, to deal with how children make intentional mistakes such as adding full stops at the end of each phrase despite no full stops being shown to them. However, we feel that the refined classification presented here provides an effective underpinning for this and such a future development will not only be a useful analytic tool, but also move us closer to our long-term diagnostic goal of differentiating individuals from the pattern of typing error they make.

9. REFERENCES

[1] Cooper, W. E. (ed.). *Cognitive Aspects of Skilled Typewriting*. Springer-Verlag, New York, 1983.

[2] Dvorak, A., Merrick, N. L., Dealey, W. L., and Ford, G. C. *Typewriting Behaviour*. American Book Company, New York, 1936.

[3] Gentner, D. R., Grudin, J. T., Larochelle, S., Norman, D. A., and Rumelhart, D. E. A glossary of terms including a classification of typing errors. In W. E. Cooper (ed.), *Cognitive aspects of skilled typewriting*. Springer-Verlag, New York, 1983, 39-43.

[4] Grudin, J. T. Error patterns in novice and skilled transciption typing. In W. E. Cooper (ed.), *Cognitive aspects of skilled typewriting*. Springer-Verlag, New York, 1984, 121-139.

[5] Kano, A., Read, J. C., and Dix, A. Children's phrase set for text input method evaluation. In *Proceedings of the fourth Nordic conference on Human-computer interaction (NordiCHI 2006)*. ACM Press, New York, NY, 2006, 449-452.

[6] Lessenberry, D. D. *Analysis of Errors*. Syracuse, NY, 1928. [Reprinted in Dvorak et al. *Typewriting Behaviour*, American Book Company, New York, 1936].

[7] MacKenzie, I. S. KSPC (keystrokes per character) as a characteristic of text entry techniques. In *Proceedings of the fourth international symposium on Human computer interaction with mobile devices*. Springer-Verlag, Heidelberg, 2001, 405-416.

[8] MacKenzie, I. S. Introduction to this special issue on text entry for mobile computing. *Human-Computer Interaction*, 17, 2/3 (2002), 141-145.

[9] MacKenzie, I. S. *Teps, A Shortened Name for the Phrase Set.* Personal Communication, 2006.

[10] MacKenzie, I. S., and Soukoreff, R. W. Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction*, 17, 2/3 (2002), 147-198.

[11] MacKenzie, I. S., and Soukoreff, R. W. Phrase sets for evaluating text entry techniques. In *CHI 2003: Extended abstracts of the SIGCHI conference on Human factors in computing systems*. ACM Press, New York, NY, 2003, 754-755.

[12] MacKenzie, I. S., and Tanaka-Ishii, K. (eds.). *Text Entry Systems: Mobility, Accessibility, Universality*. Morgan Kaufmann Publishers, San Francisco, 2007.

[13] Norman, D. A. *Categorization of action slips. Psychological Review, 88,* 1 (1981), 1-15.

[14] Read, J. C., and Horton, M. *When teenagers ttype*. In *Proceedings of British HCI 2006*. London, UK, 2006.

[15] Read, J. C., MacFarlane, S. J., and Casey, C. *Measuring the usability of text input methods for children*. In *People and Computers XV (Proceedings of HCI 2001)*. Springer, London, 2001, 559-572.

[16] Smith, H. H. Problems in diagnostic testing and remedial teaching as applied to typewriting. In *Second yearbook*. Commercial Education Association of New York City and Vicinity, New York, 1932, 53-68.

[17] Soukoreff, R. W., and MacKenzie, I. S. Measuring errors in text entry tasks: An application of the Levenshtein string distance statistic. In *CHI 2001: Extended abstracts of the SIGCHI conference on Human factors in computing systems*. ACM Press, New York, NY, 2001, 319-320.

[18] Soukoreff, R. W., and MacKenzie, I. S. *Metrics for text entry research: An evaluation of MSD and KSPD, and a new unified error metric*. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI 2003)*. ACM Press, New York, NY, 2003, 113-120.

[19] White, W. T. *Typing for Accuracy*. H. M. Rowe Company, Balitimore and Chicago. [Reprinted in Dvorak et al. *Typewriting Behaviour*, American Book Company, New York, 1936].

[20] Wobbrock, J. O., and Myers, B. A. Analyzing the input stream for character-level errors in unconstrained text entry evaluations. *ACM Transactions on Computer-Human Interaction, 13*, 4 (2006), 458-489.