



## UvA-DARE (Digital Academic Repository)

### Tracing the Development of Typewriting Skills in an Adaptive E-Learning Environment

van den Bergh, M.; Schmittmann, V.D.; Hofman, A.D.; van der Maas, H.L.J.

**DOI**

[10.2466/23.25.PMS.121c26x6](https://doi.org/10.2466/23.25.PMS.121c26x6)

**Publication date**

2015

**Document Version**

Final published version

**Published in**

Perceptual and Motor Skills

**License**

Article 25fa Dutch Copyright Act

[Link to publication](#)

**Citation for published version (APA):**

van den Bergh, M., Schmittmann, V. D., Hofman, A. D., & van der Maas, H. L. J. (2015). Tracing the Development of Typewriting Skills in an Adaptive E-Learning Environment. *Perceptual and Motor Skills*, 121(3), 727-745. <https://doi.org/10.2466/23.25.PMS.121c26x6>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*

## TRACING THE DEVELOPMENT OF TYPEWRITING SKILLS IN AN ADAPTIVE E-LEARNING ENVIRONMENT<sup>1</sup>

MATTIS VAN DEN BERGH AND VERENA D. SCHMITTMANN

*Tilburg University*

ABE D. HOFMAN AND HAN L. J. VAN DER MAAS

*University of Amsterdam*

*Summary.*—Typewriting studies which compare novice and expert typists have suggested that highly trained typing skills involve cognitive process with an *inner* and *outer loop*, which regulate keystrokes and words, respectively. The present study investigates these loops longitudinally, using multi-level modeling of 1,091,707 keystroke latencies from 62 children (*M* age = 12.6 yr.) following an online typing course. Using finger movement repetition as indicator of the inner loop and words typed as indicator of the outer loop, practicing keystroke latencies resulted in different developmental curves for each loop. Moreover, based on plateaus in the developmental curves, the inner loop seemed to require less practice to develop than the outer loop.

In order to become a skilled typist, one must master a wide variety of motor and cognitive processes, ranging from hand and finger movements to language generation and comprehension (Shaffer, 1976; Rumelhart & Norman, 1982; Salthouse, 1986; John, 1996; Wu & Liu, 2008). Older typing studies primarily focused on developing motor skills. For instance, Swift (1904) measured typewriting skills as number of words typed per hour, and Lashley (1951) focused on optimizing successive keystrokes as a function of speed and accuracy. More recent typing development studies also account for cognitive skill. Logan and Crump (2011) made an explicit distinction between motor-oriented skills and cognitive-based skills, labeled as the inner and outer loops, respectively.

The inner and outer loops are nested feedback loops that serve distinct purposes. The inner loop monitors the immediate goals; e.g., press key T by moving one finger, then press H by moving a second finger, and finally press E with another finger. The outer loop monitors the broader, semantic goals; e.g., what word or sentence is to be typed next.

This two-loop theory for typewriting is supported by several experiments (e.g., Logan & Zbrodoff, 1998; Logan, 2003; Crump & Logan, 2010a, 2010b). For example, Logan and Zbrodoff (1998) showed with a typewrit-

---

<sup>1</sup>Address correspondence to Mattis van den Bergh, Department of Methodology and Statistics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands or e-mail (m.vdnbergh@uvt.nl).

ten Stroop task that congruency of the color and the word to be typed affected response times but not the inter-keystroke interval. Hence, the meaning of the word to be typed (outer loop) is influenced by the congruency, but not by the execution of the keystrokes within a word (inner loop). In another study, Logan and Crump (2009) limited the characters to be typed to those that should be typed with one of the hands. This restriction on the inner loop resulted in an increase in errors and decrease in speed. Furthermore, they concluded that the inner loop is largely an unconscious process. For a comprehensive overview of most experiments on the inner and outer loops, see Logan and Crump (2011).

Many studies have provided support for the inner and outer loops by using an expert–novice paradigm; aspects of expert typewriting are compared to those of novices. However, motor and cognitive processes of novice and expert typists are likely to be qualitatively different, as novice typists have developed neither the inner loop nor the outer loop. Furthermore, such a cross-sectional design is of limited use for studying how typewriting develops, or more specifically how the inner and outer loops develop. Instead, a longitudinal approach is more suitable for showing developmental trends of both loops, as the same persons are measured on multiple occasions and hence within-participants differences are also assessed.

Novice typists have to acquire the more motor-oriented inner loop as well as the more cognitive-oriented outer loop. The development of the two loops can only be indirectly inferred from differences in latencies between keystrokes. Novices do not know the layout of the keyboard yet. Hence, they have to search the keyboard for each individual character. If they have to type the same character consecutively, the latency will be smaller, as this character will be typed with the same finger. The difference between the latencies of keystrokes with and without finger repetition can be interpreted as knowledge of the keyboard. Hence, the development of the inner loop (or aspects of it) can be inferred from differences in these keystroke latencies.

The outer loop relates to the meaning of the word or sentence to be typed. Outer loop development can be inferred from differences in inter-keystrokes latencies between typing words and non-words. Indeed, empirical studies show that expert typists with a developed outer loop type words faster than non-words, while novice typists type words and non-words at the same pace (Fendrick, 1937; Shaffer & Hardwick, 1968; Gentner, Larochelle, & Grudin, 1988).

All studies thus far have aggregated latencies beyond individual keystrokes, for instance by measuring the average number of words typed in a certain time interval, the time used for typing words or non-words, or the time used to type a specific number of characters. Such aggregation

does not account for parts of the observed variance and thus may considerably bias conclusions (Burstein, 1980). Therefore, this study takes a more statistically sound approach and uses a hierarchical linear model to distinguish within- and between-participants variance components. This is required to test the development of the inner and outer loops (compare with Schwartz & Stone, 1998).

It has been shown consistently that typewriting speed increases with practice (e.g., Hill, Rejall, & Thorndike, 1913) and that typing requires at least two different feedback loops. However, the way in which these loops develop has not been investigated, despite the implications for acquiring typewriting skills. Therefore, the present study traces inner and outer loop development through keystroke latencies in novice typists and assesses the contribution of both loops to overall typing speed development. Additionally, the study will investigate how much practice is necessary to reach (at least temporarily) a plateau in the development of each loop (compare with Buitrago, Schulz, Dichgans, & Luft, 2004; Maniar, Council, Prasad, Prasad, Chu, & Damiano, 2005) and, hence, which loop requires the least practice.

## METHOD

### *Participants*

Participants were selected for the study from an online typing course (Typegarden;  $N=1,226$ ). To ensure that all participants had sufficient practice, only those who had made more than 10,000 keystrokes during the course were selected. Further, to ensure that all participants had progressed sufficiently, only those who scored at the end of the course within the top 25% of all participants in terms of speed and accuracy were selected. This combination of criteria resulted in a selection of 24 boys and 38 girls with a mean age of 12.6 yr. ( $SD=1.6$ ), who in total had 1,091,707 keystroke latencies, with an average of 17,608 keystrokes per participant (range: 10,067–19,999). Only correct keystrokes that were also preceded by a correct keystroke were taken into account. This prevented confounding effects like post-error slowing. The participants agreed with the use of the anonymized data for scientific research when they took a subscription to the Typegarden system.

### *Measures*

Data for this research were obtained through Typegarden, an adaptive e-learning environment that teaches children to touch type. It has eight levels in which keys are introduced progressively. This study used only the first level, which has 270 non-words and 80 words that use the eight keys of the central row of a QWERTY-keyboard (asdfjkl;). This is the simplest level since the fingers do not have to travel over the keyboard. Each item is a letter string consisting of one or more words or non-words

shown on screen, and feedback is given by highlighting each letter as it is typed (green for correct, red for error). Each item is to be completed within 20 sec. An item's score depends on both the speed and accuracy of the response. Though the scoring rule has not been evaluated for Typegarden, a similar program called Math Garden has shown it to have excellent psychometric properties (Maris & van der Maas, 2012). Students progress at their own pace, as Typegarden is a computer adaptive program where the difficulty of the next letter string (item) is matched to the participant's current ability (Klinkenberg, Straatemeier, & van der Maas, 2011). Hence, a novice typist will receive mainly easy items, while an expert receives mainly difficult items, which causes participants to practice with items that differ (in the frequency that they are presented). Therefore, two random participants might have the same number of errors, but not the same typing skills. Once a certain level is reached, a student has the opportunity to proceed to the next level. Hence, typewriting skills do not have to be fully developed for a student to progress, as the typewriting skills can still be improved in the next level. The students practiced typing at school, but also had the option of practicing individually at home. Frequent practice was rewarded with digital coins, and 60% of the selected students practiced every other day.

### Analysis

The times between individual keystrokes varied greatly, ranging from 21 to 3,999 msec. (keystrokes outside this range were regarded as outliers and have been removed), with a mean of 656.68 msec. ( $SD = 559.22$ ). As the data were positively skewed, a natural log transformation was applied to the keystroke latencies to normalize their distribution (Fig. 1).

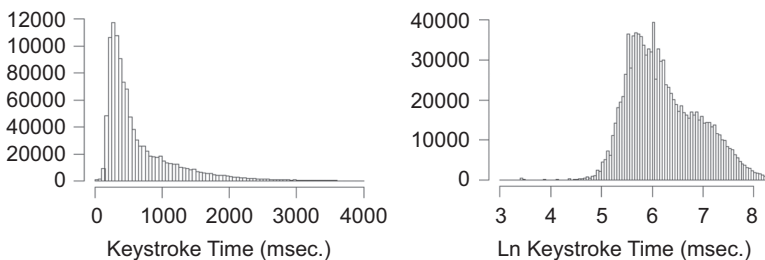


FIG. 1. The distribution of the reaction times of keystrokes

It is well known that even a small distraction can cause the reaction time of a person's individual keystrokes to lengthen (e.g., Strayer & Johnston, 2001). A more reliable measure can be extracted if the data are

grouped in fixed segments of 100 successive keystrokes. This way, the number of segments represents the amount of practice. One hundred keystrokes per second was relatively arbitrary, as a segment could also comprise, for instance, one item (a string of keystrokes) or one login session. A number of segment sizes were tested; the authors are convinced that the main results are not influenced by this choice.

The data originate from a complex sample in which observations are nested within individuals. Therefore, a distinction can be made between the variance between participants and the variance between keystrokes within participants. The variance between participants indicates differences in the participants' average successive keystroke times per segment. The variance within participants indicates the difference between keystroke times of different keystrokes for an individual participant in a specific segment. The ratio of the within-participant and between-participant variances per segment is indicative for how well a distinction can be made between participants for a given segment. Generally, there are two indices which are sensitive for this distinction: intraclass correlation (ICC) and reliability (compare with Brennan, 2000).

The change in ICC is shown in the top panel of Fig. 2. This figure indicates how necessary a multi-level model is (Hox, 2010, p.15). The ICC changes during learning and ranges from .04 in the beginning to .12 in the middle and .04 at the end. Note that small ICC's, or small differences in ICC, can indicate large differences between different typists (compare with Snijders & Bosker, 1999), and can have a great effect on the significance of related parameters (Goldstein, 2011). The bottom panel of Fig. 2 shows the reliability estimates of differences in typing speed of individuals over time. The reliability ranges from .80 to .94, and is on average .90.

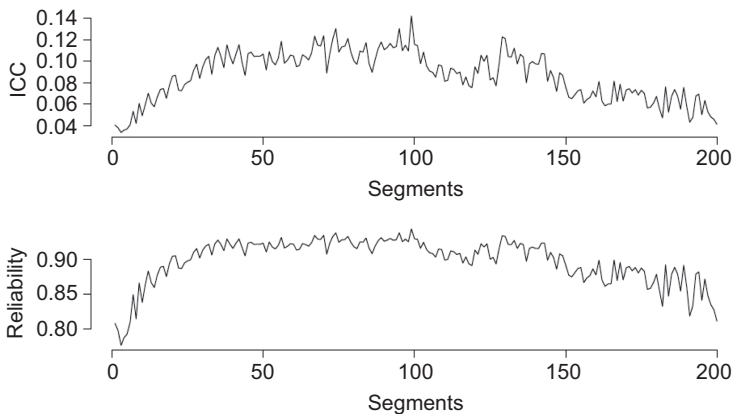


FIG. 2. ICC and reliability over segments of 100 keystrokes

### Construction of Models

The development of overall typing skills was modeled by fitting subsequent polynomial functions to the data. That is, differences in individuals' keystroke latencies were modeled as a function of powers of practice (i.e., segments of 100 keystrokes). Such polynomials are very flexible functions that can take almost any shape (depending on the order of the polynomial and the value of the individual coefficients).

If  $y_{ij}$  is the latency on the  $i$ th segment (the amount of practice in  $i$  times 100 keystrokes) of the  $j$ th individual, then a polynomial can be written as:  $y_{ij} = f_j(\text{practice}_{ij})$ . This function can be written as a regression model, which assumes that the latencies depend on powers of segment:

$$y_{ij} = \beta_{0j} + \beta_{1j} * \text{Seg}_{ij}^1 + \beta_{2j} * \text{Seg}_{ij}^2 + \dots$$

As the estimated latencies will never correspond perfectly to the observed latencies, usually the difference is taken into account by an error term. In this case, however, the error term might also depend on practice (e.g., the reliability estimates in Fig. 3b). Therefore, an individual's residuals must be modeled as a function of practice.

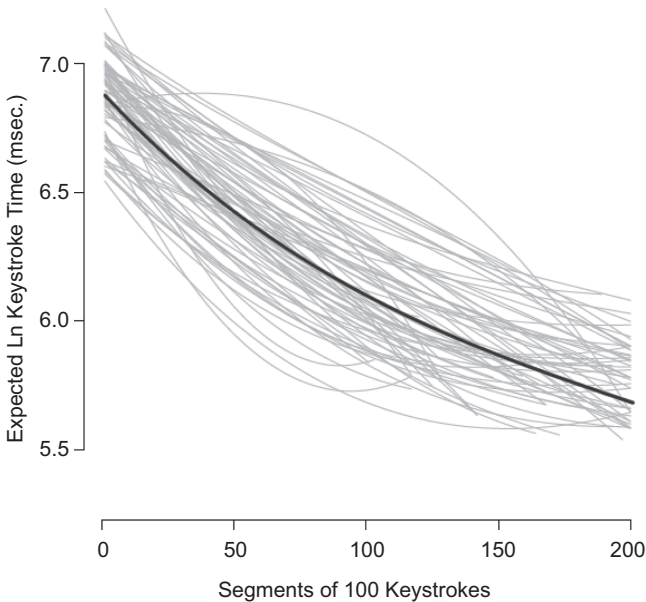


FIG. 3. The general model of average and individual development of keystroke time over segments

The individual regression coefficients ( $\beta_{0j}, \beta_{1j}, \beta_{2j}, \dots$ ) can be written as deviations from an average of the respective parameter:

$$\beta_{0j} = \beta_0 + u_{0j} \quad \beta_{1j} = \beta_1 + u_{1j} \quad \beta_{2j} = \beta_2 + u_{2j} \dots$$

For instance, a second order polynomial can be written as:

$$\begin{aligned} y_{ij} = & \beta_0 + \beta_1 * Seg_{ij}^1 + \beta_2 * Seg_{ij}^2 \\ & [(e_{0ij} + e_{1ij} * Seg_{ij}^1 + e_{2ij} * Seg_{ij}^2) \\ & (u_{00j} + u_{10j} * Seg_{ij}^1 + u_{20j} * Seg_{ij}^2)]. \end{aligned} \quad [1]$$

The model, as shown in Eq. 1, consists of a fixed part and a random part (between square brackets). The fixed part estimates the average change with practice. The first fixed parameter ( $\beta_0$ ) represents the average keystroke time for segment 0 (also known as an intercept), the second fixed parameter ( $\beta_1$ ) represents the change in keystroke time per segment, and the third fixed parameter ( $\beta_2$ ) indicates the extent that the change in keystroke time per segment changes per segment squared. The random part of the model distinguishes between deviations from the average for individuals ( $u$ 's) and deviations of the observations from the individual curves ( $e$ 's). Hence,  $e_{0ij}$  represents the deviation of the average keystroke latency of the  $j$ th individual. As the within-individuals variance might depend on practice, heteroscedasticity is modeled in terms of the polynomial. It is assumed that all residuals are normally distributed, with an expected value of 0 and a variance of  $S^2_{e_{0ij}}, \dots, S^2_{u_{20j}}$ , respectively. Furthermore, it is assumed that the residuals within and between individuals are uncorrelated ( $r_{e,u} = 0$ ).

From Equation 1, the variances within and between individuals are a function of segment. The variance within participants can be approximated as:

$$\begin{aligned} Var(\text{within} | Seg = T) = & \\ & S^2_{e_{0ij}} + 2 * Cov(e_{0ij}, e_{1ij}) * T + S^2_{e_{1ij}} * T^2 + 2 * Cov(e_{0ij}, e_{2ij}) * T^2 \\ & + \dots + S^2_{e_{1ij}} * T^4 \end{aligned} \quad [2]$$

The variance between individuals can be approximated in the same way (conforms with Goldstein, 1987). Hence, modeling polynomials with a multi-level model allows for accommodating heteroscedasticity of variances.

The order of the polynomial can be seen as an empirical matter. This study chose the polynomial that is most parsimonious and fits the data



best according to a likelihood ratio test for subsequent analysis (compare with van Veen, Evers-Vermeul, Sanders, & van den Bergh, 2013). The model can be extended to include variables indicative for the outer loop (words) or inner loop (finger repetition). Main effects of words or finger repetition indicate that the intercept between outer and inner loops differs from the average, whereas interactions with practice show that the development of both outer and inner loops differs from the average development.

To determine whether the inner loop develops differently from the outer loop, the inflection points (the points where the change in keystroke latencies becomes zero) will be assessed for the four possible circumstances: non-words and no finger repetition (NW-NFR), words and finger repetition (W-NFR), non-words and finger repetition (NW-FR), and words and finger repetition (W-FR). The inflection points will be determined with the first-order derivative, while the second-order derivative will indicate whether an inflection point is a minimum or a maximum. The maxima will not be of interest as they indicate the start of the development. The minima are of interest as they indicate the end of the development. As not every individual's polynomial has to have an inflection point (because not every student has to finish his development), a selection of participants with an inflection point will be made. If an inflection point of one circumstance has a lower segment number (i.e., took less practice to reach) than another circumstance, then the development of the first circumstance finished first, thereby indicating which loop finished developing first. This will be done for both the average curve of each circumstance, as well as for the individual polynomials.

## RESULTS

To describe the development in latencies, several models were fitted. Both the fixed and random parts of these models increased in complexity. The fit of each model, along with the difference in fit between consecutive models, is presented in Table 1 and expressed by  $-2 \log$  likelihoods. From the comparison between models it is apparent that a model with a fixed linear component, allowing for differences in keystroke latencies between segments, fitted better to the data than a model with only an intercept ( $\Delta\chi^2 = 237,684$ ;  $df = 1$ ;  $p < .001$ ). Allowing the linear component to vary both within and between participants improved the fit, as can be seen in Rows 3 and 4 of Table 1 ( $\Delta\chi^2 = 23,061$ ;  $df = 4$ ;  $p < .001$ ). The fit of consecutive models increased up to the third order polynomial  $\beta_3 * Seg_{ij}^3$ . The third order term is allowed to differ within individuals ( $Se_{3ij} * Seg_{ij}^3$ ) and between individuals ( $Su_{30j} * Seg_{ij}^3$ ). As shown in Table 1, a fixed quartic term  $\beta_4 * Seg_{ij}^4$  did not improve the model fit ( $\Delta\chi^2 = 2.00$ ;  $df = 1$ ;  $p < .84$ ).

TABLE 1  
LIKELIHOOD RATIO TEST OF THE GENERAL MODEL

Model	No. Parameters	-2LL	$\Delta\chi^2$	$\Delta df$	<i>p</i>
$y_{ij} = \beta_{0ij} + [e_{0ij} + u_{00j}]$	3	2,255,410			
$+ \beta_1 * Seg_{ij}^1$	4	2,017,726	237,684	1	<.001
$+ e_{1ij} * Seg_{ij}^1$	6	2,009,725	8,001	2	<.001
$+ u_{10j} * Seg_{ij}^1$	8	1,994,665	15,060	2	<.001
$+ \beta_2 * Seg_{ij}^2$	9	1,986,498	8,167	1	<.001
$+ e_{2ij} * Seg_{ij}^2$	12	1,986,081	417	3	<.001
$+ u_{20j} * Seg_{ij}^2$	15	1,978,324	7,757	3	<.001
$+ \beta_3 * Seg_{ij}^3$	16	1,978,144	180	1	<.001
$+ e_{3ij} * Seg_{ij}^3$	20	1,977,901	243	4	<.001
$+ u_{30j} * Seg_{ij}^3$	24	1,975,023	2,878	4	<.001
$+ \beta_4 * Seg_{ij}^4$	21	1,975,021	2	1	.84

Hence, a third order polynomial was necessary to describe the observed average latencies over participants and the changes in variance within and between individuals. The parameter estimates for this model are presented in Table 2. As the change per segment directly depends on the scale of the segment variable, this has been centered and runs in 200 steps from -10 to 10. The first column of Table 2 shows the parameter estimates for the model of general development. It can be seen that the average time between keystrokes decreases significantly over (the recoded) segments. The average log transformed keystroke time at Segment 100 (keystroke 9,900 until keystroke 10,000; the intercept) is estimated as 6.09 (441 msec.), and changes continuously by -0.057 per segment. As segments have been recoded from -10 to 10 in steps of 0.1, this amounts to a change of  $-0.057 * 0.1 = -0.0057$  per segment. Simultaneously, there is also an increase with  $0.504 * 10^{-3}$  per squared segment and a decrease with  $-0.178 * 10^{-3}$  for the cubed segment.<sup>2</sup> Therefore, the expected log transformed successive keystroke time in the first observed segment (with a recoded value of -10) is  $[6.088 - (0.057 * -10) + (0.504 * (-10)^2 * 10^{-3}) + (-0.178 * (-10)^3 * 10^{-3})] = 6.786$  (885 msec.), while the expected log transformed successive keystroke time for the final segment (10) is estimated as 5.390 (219 msec.). Hence, the average difference in keystroke times between the first and last segment is  $[e^{6.786} - e^{5.390}] = 666$  msec. In Fig. 3, the average keystroke time is presented by means of a black solid line.

In this general model, both the fixed parameters and the random parameters have been estimated. The random parameters show differences

<sup>2</sup>This shows one of the reasons for recoding the segments. Without recoding, the quadratic and cubic parameters would have been even smaller.

TABLE 2  
 FIXED AND RANDOM PARAMETERS OF MODELS TO DESCRIBE THE LEARNING OF TOUCH TYPING  
 (FOR CONVENIENCE, SEGMENT HAS BEEN RECORDED FROM -10 TO 10)

	General Development		Word & Finger Repetition	
	Est.	SE	Est.	SE
Fixed Part				
Seg <sup>0</sup>	6.088	0.015	6.152	0.016
Seg <sup>1</sup>	-0.057	0.003	-0.056	0.003
Seg <sup>2</sup> *10 <sup>3</sup>	0.504	0.259	0.141	0.582
Seg <sup>3</sup> *10 <sup>3</sup>	-0.178	0.035	-0.324	0.058
Word			-0.120	0.002
W*Seg <sup>1</sup> *10 <sup>3</sup>			-0.328	0.546
W*Seg <sup>2</sup> *10 <sup>3</sup>			0.646	0.046
W*Seg <sup>3</sup> *10 <sup>3</sup>			-0.071	0.009
FR			-0.472	0.003
FR*Seg <sup>1</sup>			0.031	0.001
FR*Seg <sup>2</sup> *10 <sup>3</sup>			-1.635	0.055
FR*Seg <sup>3</sup> *10 <sup>3</sup>			-0.074	0.011
W*FR			0.139	0.004
W*FR*Seg <sup>1</sup>			0.002	0.001
Random Part <sup>a</sup>				
Variance within individuals				
S <sup>2</sup> <sub>e<sub>0ij</sub></sub>	0.342		0.321	
S <sup>2</sup> <sub>e<sub>1ij</sub></sub>	0.001		0.001	
S <sup>2</sup> <sub>e<sub>2ij</sub></sub> *10 <sup>3</sup>	0.013		0.003	
S <sup>2</sup> <sub>e<sub>3ij</sub></sub> *10 <sup>3</sup>	<0.001		<0.001	
Variance between individuals				
S <sup>2</sup> <sub>μ<sub>0ij</sub></sub>	0.047		0.045	
S <sup>2</sup> <sub>μ<sub>1ij</sub></sub>	0.001		0.001	
S <sup>2</sup> <sub>μ<sub>2ij</sub></sub> *10 <sup>3</sup>	0.024		0.032	
S <sup>2</sup> <sub>μ<sub>3ij</sub></sub> *10 <sup>3</sup>	<0.001		<0.001	

<sup>a</sup>Covariances of the general model are presented in the Appendix.

between participants. The variance of differences between individuals at the intercept, for instance, is estimated as 0.05 (see Table 2). Hence, an 80% confidence interval for differences between individuals of segment 0 (i.e., the 100th segment from the start) ranges from 5.73 to 6.37. And the average linear change per segment equals -0.06, but this change differs be-

tween participants (variance=0.001). Thus, an 80% confidence interval of the differences between participants for the linear change with segment ranges from  $-0.10$  to  $-0.02$ . That is, for some participants the linear decrease in keystroke time is steeper than for others. The same holds for the quadratic coefficient ( $80\%CI = -4.11 * 10^{-3}, 5.12 * 10^{-3}$ ) and the cubic coefficient ( $80\%CI = -0.73 * 10^{-3}, 0.38 * 10^{-3}$ ).

These random terms can be used to approximate the variance in each segment (Equation 2) as well as the variance within and between participants. For instance, at the intercept the variance within individuals is estimated as 0.34. For the first and last segments, the variance within individuals is approximated as 0.88 and 0.55, respectively. Thus, the difference between keystrokes within participants clearly decreases with practice.

The variance between participants at the intercept is estimated as 0.05. The variance between participants increases significantly with practice; at the first segment this variance is estimated as 0.12, whereas at the last segment the estimate is 0.90. In Fig. 3, the grey lines represent the estimated polynomials for the individual participants. The average change in keystroke time (on the log scale) is presented by a black solid line.

#### *Words and Finger Repetition*

In the next analysis, effects of (non-)words and (no) finger repetition on development of typing skills were compared. A likelihood ratio test showed that the model with a three-way interaction between the linear term, finger repetition, and words provides the best fit at an  $\alpha$  level of .05 (Table 3).

TABLE 3  
LIKELIHOOD RATIO TEST OF THE WORD AND FINGER REPETITION (FR) EFFECT

Model	No. Parameters	-2LL	$\Delta \chi^2$	$\Delta df$	<i>p</i>
General model	20	1,975,021			
+Words	21	1,972,762	2,215	1	< .001
+Words * $Seg_{ij}^1$	22	1,972,399	400	1	< .001
+Words * $Seg_{ij}^2$	23	1,972,174	273	1	< .001
+Words * $Seg_{ij}^3$	24	1,972,037	118	1	< .001
+FR	25	1,879,242	92,701	1	< .001
+FR * $Seg_{ij}^1$	26	1,861,802	17,473	1	< .001
+FR * $Seg_{ij}^2$	27	1,860,730	1,075	1	< .001
+FR * $Seg_{ij}^3$	28	1,860,663	64	1	< .001
+FR * Word	29	1,859,542	1,129	1	< .001
+FR * Words * $Seg_{ij}^1$	30	1,859,537	4	1	.04
+FR * Words * $Seg_{ij}^2$	31	1,859,537	1	1	.31

The estimates of the coefficients of the final model are displayed in Table 2. The main effects of words and finger repetition were significant, as well as the interaction between words and finger repetition and the interaction between words, finger repetition, and segment. Hence, the change in latencies with practice when typing words differed from the latencies when typing non-words. For instance, at the intercept words were typed faster than non-words ( $-0.12$ ). At the end of the study, the average latency in keystroke time in milliseconds for words was  $[6.152 - 0.120 + (-0.056 + -0.328 * 10^{-3}) * 10 + (0.141 * 10^{-3} + 0.646 * 10^{-3}) * 10^2 + (-0.324 * 10^{-3} + -0.071 * 10^{-3}) * 10^3] = 5.152$ , whereas the average latency for non-words was 5.282.

At the intercept, the difference in average latency due to finger repetition was  $-0.47$ , indicating that keystrokes involving finger repetition were faster. At the end of the study, the average latency for items with finger repetition was  $[6.152 - 0.472 + (-0.056 + -0.031) * 10 + (0.141 * 10^{-3} + 1.635 * 10^{-3}) * 10^2 + (-0.324 * 10^{-3} + -0.074 * 10^{-3}) * 10^3] = 4.590$ , whereas the average for items without finger repetition equaled 5.282. Note that this number is the same as the previously mentioned average latency for non-words, because these non-word latencies were estimated for when there was no finger repetition. Hence, 5.282 reflected the average latency for non-words with no finger repetition.

Both the effects for words and finger repetition changed with practice. Figure 4 shows for all four combinations of words and finger repetition the average change with practice. Figure 5 shows that without finger repetition the estimated average (ln) keystroke latency initially hardly differed between non-words and words. However, with practice a difference emerged: without finger repetition, words were typed faster than non-words. When finger repetition was present, there was initially a difference between non-words and words, with non-words showing smaller latencies than words. This was probably due to the presence of a set of very

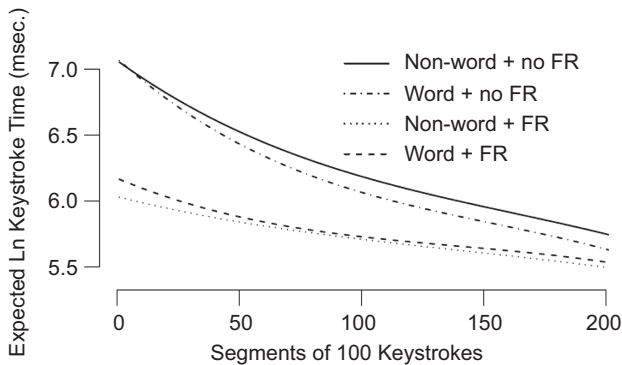


FIG. 4. Estimations of the combinations of word and finger repetition (FR) effects

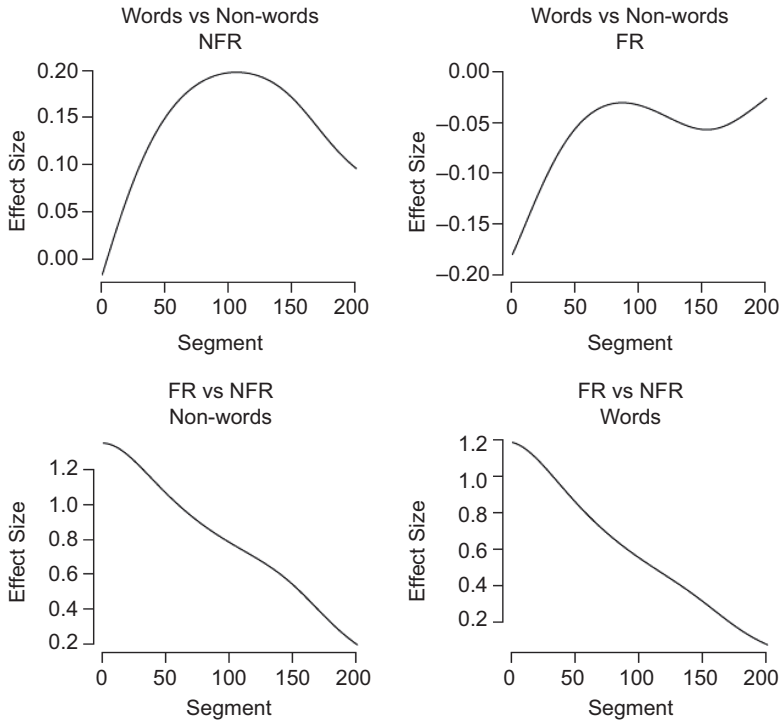


FIG. 5. Effect sizes of finger repetition and words (Cohen's  $d$ )

easy items with many repetitions (such as “fff,” “aass,” “aaasssdddfff”). The effect of finger repetition existed, since items with finger repetition were typed faster than items without finger repetition. However, this difference between words and non-words was reduced with practice.

Because the sample of this study was large, the power of the significance tests was large. Hence, it is possible that some significant results were only due to very small differences in latencies. Effect sizes show whether the assessed differences were substantial or negligible. Cohen's  $d$  (Cohen, 1988) was computed for each segment.<sup>3</sup> The initial effect size for the typing of words without finger repetition equaled  $-0.02$  and increased to a maximum of  $0.20$  at the 112th segment. Thereafter, the effect size decreased slightly to  $0.15$  at the final segment (see the top left part of Fig. 5). This indicates that after some practice words were typed faster than non-words, but this effect was small at best.

<sup>3</sup>Effect sizes over  $0.8$  are considered large, over  $0.5$  as medium, and over  $0.2$  as small.

The total effect size for the word effect when finger repetition was present (see the top right part of Fig. 5) had an initial value of  $-0.19$ , indicating that in the beginning words were typed slower than non-words. It increased slightly, but can be considered as very small. The initial effect size for finger repetition when non-words were typed was  $1.43$ . This decreased quickly and had a value of  $0.32$  at the final segment (see the bottom left part of Fig. 5). The effect size for finger repetition when words were typed showed the same pattern of almost linear decrease, but with smaller absolute values. It started at  $1.26$  and ended at  $0.12$  (see the bottom right part of Fig. 5).

### *Inflection Points*

To assess the moment at which development (at least temporarily) came to a halt, the inflection points of the curves were determined. Such inflection points indicate the segment, or the amount of practice, at which no change in keystroke latencies are expected and a minimum occurs. The average development, as shown in Figs. 3 and 4, shows a continuous decrease in inter-keystroke latencies. Hence, the average development does not show any inflection point; on average the children had not finished developing their skills for typing words and non-words, with and without finger repetition.

Although the average change over time does not show an inflection point, this does not necessarily hold for every individual curve. Comparison of how much practice produces an inflection point in each circumstance (NW–NFR, W–NFR, NW–FR, and W–FR) for every individual allows determination of which loop developed faster. Note that a comparison can only be made if an individual's data actually has an inflection point in both circumstances. If there is no inflection point in a certain circumstance, the development for that circumstance has not been (at least temporarily) finished.

In Table 4, the last column shows the total number of inflection points in each of the circumstances. For NW–NFR there were five participants who showed an inflection point, while there were four participants with an inflection point for W–NFR. For the circumstances with finger repetition, NW–FR and W–FR, there were 34 and 31 participants with an inflection point, respectively. There were more participants with an inflection point in the circumstances with finger repetition than those without. This shows that the required practice for development of finger repetition was less than the development of no finger repetition. The effect of words was less obvious. The number of participants with an inflection point with non-words was consistently higher than with words, but the differences were small.

Furthermore, Table 4 also shows whether the (temporary) plateau in one of the circumstances preceded that of another circumstance (above

TABLE 4  
 TOTAL NUMBER OF PARTICIPANTS WITH (AT LEAST TEMPORARILY) A FINISHED DEVELOPMENT  
 IN THE DIFFERENT CONDITIONS AND THE NUMBER OF PARTICIPANTS REQUIRING MORE  
 AND LESS PRACTICE IN THE DIFFERENT CONDITIONS TO REACH A (TEMPORARY)  
 PLATEAU (LESS ABOVE AND MORE BELOW THE DIAGONAL)

Circumstance	NW-NFR		W-NFR		NW-FR		W-FR		Total
	Total	Change	Total	Change	Total	Change	Total	Change	
NW-NFR			1	4	5	5	5	5	5
W-NFR	2	4			4	4	4	4	4
NW-FR	0	5	0	4			12	27	34
W-FR	0	6	0	4	14	27			31

the diagonal) or whether the (temporary) plateau occurred later than in another circumstance (below the diagonal). For instance, the first row shows that for one out of four cases the development in the NW-NFR circumstance required less practice than in the word combined with no finger repetition. For five (out of five) participants the NW-FR and W-FR circumstances required less practice to reach a (temporary) plateau than NW-NFR.

The first column of Table 4 shows that for two of four participants W-NFR required more practice than NW-NFR. Because the first row already indicated that one participant required less practice, there is one participant who needed the same amount of practice in both circumstances. No participants (out of five) required more practice to reach a (temporary) plateau in NW-FR and W-FR compared to NW-NFR.

The last two cells of the second row and the second column indicate that all four participants with an inflection point in W-NFR and NW-FR or W-FR required less practice in the circumstances with finger repetition. The final comparison between NW-FR and W-FR in the final cell of the last row and column indicated that 12 participants (out of 27) required less practice to reach a (temporary) plateau in the W-FR circumstance, while 14 participants required more practice in the W-FR circumstance. Also, between these circumstances there is one participant who needed similar practice in both circumstances to reach a (temporary) plateau. Hence, the relationship between words (compared to non-words) and typing development does not seem to be straightforward.

#### DISCUSSION

This study investigated the typewriting skill development in an adaptive learning environment. The authors analyzed data from the first game of an online course on typewriting in which the eight characters of the home row were learned. Data were collected at the keystroke level; therefore, the number of observations was enormous. Since keystroke latencies



were nested within learners, development was modeled using a multi-level approach.

This multi-level model was relatively parsimonious, with four fixed parameters for the amount of practice and seven random parameters for the differences between and within individuals. It was shown that the average keystroke latency decreased with practice, but that the learning curves of each individual differed notably. Based on analyzing the inflection points in each individual's polynomial, not all participants had reached their minimal keystroke latencies yet. Hence, their development in this task was not yet completed. This is not surprising, as after reaching a certain level the second game becomes accessible and it is an individual's own choice to continue the first game or to start playing the second game with more letters of the keyboard.

Both the inner loop, indicated by the decreasing finger repetition effect, and the outer loop, indicated by the word effect, appeared to emerge with practice (this conforms with Logan & Crump, 2011). The loops developed differently, and both effects contributed significantly to the model of overall typing development. In general there was no plateau in development (a vanishing rate of change in keystroke latencies), but these plateaus were found in some individual developmental curves. Comparing the amount of practice needed to finish development between the different circumstances of (non-)words and (no) finger repetition for each individual indicated that many more individuals finished their development in the circumstances with finger repetition than in the circumstances without. The results for the (non-)words are not so clear, since some individuals finished developing faster with words while others finished developing faster with non-words. This also made a comparison between the words and finger repetition more difficult, but the strong effect of finger repetition on the development compared to the ambiguous effect of words indicated that the development of the inner loop is finished before the development of the outer loop. This is in concordance with previous findings that while the associations between keys and finger movements are helpful for basic typing, associations between words and letters are required for skilled typing (Yamaguchi & Logan, 2014).

In the present study, the development of average keystroke latencies was analyzed per 100 keystrokes. Such an analysis neglects the natural boundaries between items, which were words or non-words with different number of characters. The proposed model can be expanded to a so-called cross-classified model (Goldstein, 2011) in which both the variance between participants and the variance between items are estimated simultaneously. This allows for a more precise analysis of item characteristics. Alas, this was not possible in the present study, as the adaptive na-

ture of the Typegarden allocated the demanded items to the ability of the participant.

Another consequence of the allocation of items was that participants with the same amount of practice did not receive the same items. However, because the presented items depended on the ability of the participants, scaffolding took place for the development of typewriting. Hence, the results should be seen as generalizable for this type of learning. The results of this paper showed how the finger repetition effect disappears and the word effect emerges, indicating the development of the inner and outer loops. The development of the inner loop seems to be finished before the outer loop, as the word effect emerges before the development of the finger repetition is finished. This suggests that the development of the inner and outer loops occur separately.

#### REFERENCES

- BRENNAN, R. L. (2000) Performance assessments from the perspective of generalizability theory. *Applied Psychological Measurement*, 24(4), 339-353.
- BUITRAGO, M. M., SCHULZ, J. B., DICHGANS, J., & LUFT, A. R. (2004) Short and long-term motor skill learning in an accelerated rotarod training paradigm. *Neurobiology of Learning and Memory*, 81(3), 211-216.
- BURSTEIN, L. (1980) The analysis of multilevel data in educational research and evaluation. *Review of Research in Education*, 158-233.
- COHEN, J. (1988) *Statistical power analysis for the behavioral sciences*. (2nd ed.) Hillsdale, NJ: Erlbaum.
- CRUMP, M. J. C., & LOGAN, G. D. (2010a) Episodic contributions to sequential control: learning from a typist's touch. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 662-672.
- CRUMP, M. J. C., & LOGAN, G. D. (2010b) Hierarchical control and skilled typing: evidence for word level control over the execution of individual keystrokes. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 36, 1369-1380.
- FENDRICK, P. (1937) Hierarchical skills in typewriting. *Journal of Educational Psychology*, 28, 609-620.
- GENTNER, D. R., LAROCHELLE, S., & GRUDIN, J. (1988) Lexical, sublexical, and peripheral effects in skilled typewriting. *Cognitive Psychology*, 20, 524-548.
- GOLDSTEIN, H. (1987) *Multilevel models in educational and social research*. London, UK: Charles Griffin.
- GOLDSTEIN, H. (2011) *Multilevel statistical models*. Vol. 922. West Sussex, UK: Wiley.
- HILL, L. B., REJALL, A. E., & THORNDIKE, E. L. (1913) Practice in the case of typewriting. *Pedagogical Seminary*, 20, 516-529.
- HOX, J. J. C. M. (2010) *Multilevel analysis: techniques and applications*. (2nd ed.) New York: Routledge.
- JOHN, B. (1996) TYPIST: a theory of performance in skilled typing. *Human-Computer Interaction*, 11, 321-355.
- KLINKENBERG, S., STRAATEMEIER, M., & VAN DER MAAS, H. L. J. (2011) Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2), 1813-1824.

- LASHLEY, K. S. (1951) The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley. Pp. 112-136.
- LOGAN, G. D. (2003) Simon-type effects: chronometric evidence for keypress schemata in typewriting. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 741-757.
- LOGAN, G. D., & CRUMP, M. J. (2009) The left hand doesn't know what the right hand is doing: the disruptive effects of attention to the hands in skilled typewriting. *Psychological Science*, 20(10), 1296-1300.
- LOGAN, G. D., & CRUMP, M. J. (2011) Hierarchical control of cognitive processes: the case for skilled typewriting. *Psychology of Learning and Motivation: Advances in Research and Theory*, 54, 1-27.
- LOGAN, G. D., & ZBRODOFF, N. J. (1998) Stroop-type interference: congruity effects in color naming with typewritten responses. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 978-992.
- MANIAR, H. S., COUNCIL, M. L., PRASAD, S. M., PRASAD, S. M., CHU, C., & DAMIANO, R. J. (2005) Comparison of skill training with robotic systems and traditional endoscopy: implications on training and adoption. *Journal of Surgical Research*, 125(1), 23-29.
- MARIS, G., & VAN DER MAAS, H. (2012) Speed-accuracy response models: scoring rules based on response time and accuracy. *Psychometrika*, 77(4), 615-633.
- RUMELHART, D. E., & NORMAN, D. A. (1982) Simulating a skilled typist: a study of skilled cognitive-motor performance. *Cognitive Science*, 6, 1-36.
- SALTHOUSE, T. A. (1986) Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin*, 99, 303-319.
- SCHWARTZ, J. E., & STONE, A. A. (1998) Strategies for analyzing ecological momentary assessment data. *Health Psychology*, 17(1), 6-16.
- SHAFFER, L. H. (1976) Intention and performance. *Psychological Review*, 83, 375-393.
- SHAFFER, L. H., & HARDWICK, J. (1968) Typing performance as a function of text. *Quarterly Journal of Experimental Psychology*, 20, 360-369.
- SNIJEDERS, T., & BOSKER, R. (1999) *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- STRAYER, D. L., & JOHNSTON, W. A. (2001) Driven to distraction: dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, 12(6), 462-466.
- SWIFT, E. J. (1904) The acquisition of skill in type-writing: a contribution to the psychology of learning. *Psychological Bulletin*, 1(9), 295-305.
- VAN VEEN, R., EVERS-VERMEUL, J., SANDERS, T., & VAN DEN BERGH, H. (2013) The influence of input on connective acquisition: a growth curve analysis of English *because* and German *weil*. *Journal of Child Language*, 40(5), 1003-1031.
- WU, C., & LIU, Y. (2008) Queuing network modeling of transcription typing. *ACM Transactions on Computer-Human Interaction*, 15, 1-45.
- YAMAGUCHI, M., & LOGAN, G. D. (2014) Pushing typists back on the learning curve: revealing chunking in skilled typewriting. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2), 592-612.

## APPENDIX A

COVARIANCE MATRICES OF THE GENERAL MODEL FOR THE BETWEEN- AND WITHIN-PARTICIPANTS  
VARIANCE COVARIANCE MATRIX BETWEEN PARTICIPANTS

	$\beta_0$	<i>SE</i>	$\beta_1$	<i>SE</i>	$\beta_2$	<i>SE</i>
$\beta_0$	0.04305	0.00773	-0.00080	0.00066	-0.00036	0.00010
$\beta_1$	-0.00080	0.00066	0.00062	0.00011	0.00006	0.00001
$\beta_2$	-0.00036	0.00010	0.00006	0.00001	0.00001	0.00000

COVARIANCE MATRIX WITHIN PARTICIPANTS

	$\beta_0$	<i>SE</i>	$\beta_1$	<i>SE</i>	$\beta_2$	<i>SE</i>	$\beta_3$	<i>SE</i>
$\beta_0$	0.34201	0.00098	-0.00399	0.00018	0.00000	0.00000	0.00000	0.00000
$\beta_1$	-0.00399	0.00018	0.00100	0.00013	0.00007	0.00001	-0.00001	0.00000
$\beta_2$	0.000000	0.00000	0.00007	0.00001	0.00000	0.00000	-0.00000	0.00000
$\beta_3$	0.000000	0.00000	-0.00001	0.00000	-0.00000	0.00000	0.00000	0.00000