Analyzing chunk pauses to measure mathematical competence: Copying equations using 'centre-click' interaction.

Peter C-H. Cheng (p.c.h.cheng@sussex.ac.uk)
Department of Informatics, University of Sussex
Brighton, BN1 9QJ, UK

Abstract

Mathematical competence can be evaluated by analyzing pauses between strokes that occurring whilst an individual copies equations. These pauses provide a temporal signal that reflects the cognitive effort to process chunks. 'Centre-click' interaction with a mouse and response grid on a computer display is introduced as new technique for measuring the temporal chunk signal. Alternative pause measures and forms of normalization are explored. It is shown that centre-click copying can be used to measure mathematical competence.

Keywords: chunks, mathematical competence, pauses analysis, graphical interface, copying equations, centre-click

Introduction

This paper continues our investigations into the measurement of complex high-level cognitive skills using simple assessments tasks. Cheng & Rojas-Anaya (2007) found that the people with substantially different levels of mathematical competence could be readily differentiated in the simple task of freehand copying mathematical formulas using a measure based on the *temporal chunk signal* manifested in the duration of pauses between pen strokes. Cheng (2014) found that both the third quartile and interquartile range of pause durations are potentially effective measures of mathematical competence in the freehand copying of equations. Beyond mathematics, Zulkifli (2013) showed that the competence of speakers of English as second language could be assessed using the third quartile of pauses in sentence coping tasks. Van Genuchen and Cheng (2009) found that the mean duration of the pauses of Dutch children writing just-memorized sentences in their native language distinguishes between children of seven and eight years of age. In general, it appears that measures based on the pauses that occur in copying and writing can be used to assess the level of competence in high-level cognitive skills.

This is feasible because of a well-establish and widely-known phenomenon: the duration of pauses between motor actions is longer for the processing elements that occur at the transition between two chunks than for element within the same chunk (e.g., van Genuchten & Cheng, 2010). For memorized chunks, the production of the first element in a chunk is slowed by the time needed retrieve and place the chunk into working memory, but once there all its elements are immediately available. In copying tasks the production of the first element is slowed by the processing associated with the perception, encoding and preparation of the chunk. Individuals with greater knowledge or more competence in a particular domain possess familiar chunks for that domain, which has two effects that underpins the temporal chunk signal that we are using to measure competence. First, familiar chunks are processed more rapidly, because their occurrence in a stimulus will be perceived as a unit and the retrieval of elements of the chunk from memory leads to rapid processing. Without such chunks stimuli elements must be perceived and encoded individually, which consumes more time. Second, because (sub-)chunks are aggregated during learning, familiar chunks may contain a relatively large number of elements that is greater than the number of elements that can be individually processed at the same time: the capacity of working memory is about four chunks for resource intense tasks (Cowen, 2001). Copying is such as task. Thus, the intra-chunk pauses will be shorter and less frequent for people familiar with a target stimulus, whereas others will have more numerous and longer pauses associated with small groups of elements due to the novelty of the stimulus. The third quartile (Q3) of individual's distributions of pauses can be used to detect these longer and more frequent pauses, so they may serve as an effective measure of competence (Cheng, 2014; Zulkifli, 2013).

The temporal chunk signal has also been used to investigate cognitive strategies in free-hand drawing (Obaidellah & Cheng, 2009; Roller & Cheng, 2014), so it might to be a good indicator of cognitive performance. However, can it be used to evaluate performance with forms of interaction more generally beyond handwriting and drawing? Pauses have been widely used to study the nature of typewriting; for instance, competence in transcription typewriting (e.g., Salthouse, 1986). However, such studies require the aggregation of data over many participants in order to find any reliable effects: this suggests measurement of individual competence using typing is unlikely to be feasible. Moreover, in a pilot experiment on the typewritten copying of English sentences (Ismail, 2010) found no significant relation between second language competence and pauses measures. Large individual differences in typing skill likely masks the temporal chunk signal.

So, is it feasible for the temporal chunk signal to be found at a level that can be used to measure competence in any forms of interaction beyond handwriting? To this end the 'centre-click' method of interaction was devised and is evaluated in the experiment here. Centre-click interaction was designed to exploit computer mouse skills, because mouse movement and clicking (button presses) are relatively simple and are uniform of across individuals, in contrast to the complexity and variability of typewriting.

Fig 1 shows the centre-click response grid used in the experiment, which is displayed on a conventional computer display. Participants copy equations by clicking on the square with the symbol they wish to input, as if it is a button. (In the present setup no visual or audio feedback is

	=	+	−	×	/	
x	*y*	*a*	*b*	*c*	*u*	*v*
0	1				2	3
4	5		■		6	7
8	9				(	)
√	∫	*dx*	—	*d*	e	ln
		2	3	x		

Fig 1. Centre-click response grid

given to the clicks). Critically, participants are trained to click the black square in the center of response grid between each symbol click. This *Return To Centre* (RTC) action aims to make the distances that the mouse moves to reach each symbol fairly uniform. (In a pilot experiment in which participants centre-click copied equations but the distance was not controlled, no correlation between mathematical competence and pauses measures was found.) The participants in the present experiment were given no directions about whether to think of RTC actions as associated with the preceding or following symbol click. To copy stimulus equations participants execute RTC and symbol clicks as required.

Symbols in the response grid are largely grouped according to related mathematical meanings, on the assumption that this may differentially assist participants who have greater mathematical competence.

Cheng (2014) showed that the third quartile (Q3) of pauses between written strokes was a good measure, where a pause is the time between the placing of the pen to the paper to inscribe the current symbol minus the time at the lifting of the pen at the end of the previous symbol. Theoretically, for the centre-click interaction several measures can be imagined depending on the assumed interaction strategy. (1) If participants do a RTC automatically after the completion of a symbol click, then chunk processing will be associated with symbol clicks and hence the pauses of interest are those just before the symbol click. The third quartile measure based on such pauses will be designated $Q3_{sym}$. (2) If participants' chunk processing occurs before the RTC, delaying the RTC until just before the symbol click, then the target pauses are associated with the RTC: designation $Q3_{RTC}$. (3) Chunk processing might sometimes occur just before RTCs or just before symbol clicks, in which case all the pauses are potentially relevant, so the pauses for both may be pooled: $Q3_{All}$. (4) Chunk processing for each symbol might consistently be distributed between both RTCs and symbols clicks, so the sum of the pauses could be considered: $Q3_{RTC+sym}$. (5) Given case (1) in which all the chunk processing occurs just before the symbol click, we might assume the RTC pauses simply reflect basic move and click times. Thus, those pauses might be used to normalize symbol click times for individual differences in such elementary actions, by subtracting RTC pauses from symbol pauses: $Q3_{sym\text{-}RTC}$.

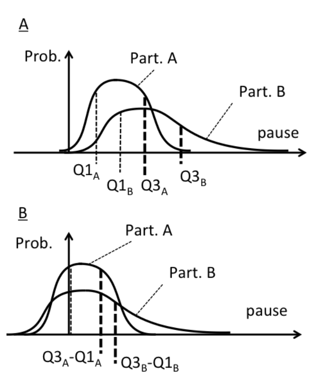

Fig 2. Elementary skills normalization.

Which of these pauses measure will best predict competence?

In addition, the experiment will also examine if normalizing for individual differences will improve the basic Q3 pause measure. For the handwritten copying of equations Cheng (2014) attempted to two forms of normalization. (a) The idea of *normalization for elementary skills* is illustrated in Fig 2. Assume that the duration of first quartile, Q1, pauses are representative of the time to execute common elementary inter-chunk processes. If these are quite different, as shown in Fig 2A, then these elementary skills will individually affect Q3 to different extents (horizontal position of the distributions). However, by subtracting Q1 from Q3, Fig 2B, the Q3 measures for both participants are put on an equal footing. Q3-Q1 is of course the interquartile range, IQR. (b) *Normalization for basic mathematical skill* gives a measure that aims to estimate pauses that are associated with handling more challenging parts of expressions, such as the perception and production non-alphanumeric symbols in non-linear arrangements, over and above processes that are required for basic components of expressions, such as simple linear sequences of numbers. Challenging aspects are likely to better differentiate people with different levels of competence. This is illustrated in Fig 3, where curve *M* is the measured distribution of pauses for a participant copying a complex equation and curve *B* is the distribution for copying a simple list of numbers. Now, if distribution B is subtracted from distribution M, a distribution for challenging aspects, *N*, is estimated. The actual normalized measure is calculated by subtracting $Q3^B$ from the measured Q3 ($Q3^N$=Q3-$Q3^B$). Will the two forms of normalization improve the basic Q3 pause measure?

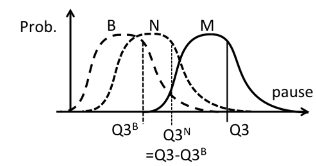

Fig 3. Basic mathematical skill normalization.

Method

The experiment was conducted by student experimenters on the *Psychological Methods for System Evaluation* module of the HCI MSc at the University of Sussex. The University's ethics committee granted ethics approval.

Participants

The 22 participants were adult friends and relatives of the students, who were deliberately recruited to span a range of mathematical experience. They were all competent computer users and had no impairments.

Materials

To independently assess their mathematical competence the participants completed the same questionnaire as used by Cheng (2014). It had three equal weight parts that assessed: (a) mathematical qualifications and current use of mathematics; (b) mathematical knowledge by posing problems with multiple choice answers; (c) participants' confidence through their ratings of the answers to the questions in (b).

E1 56789 12345 34567 54321 98765 76543 56789 23456 45678

E2 193 826 745 931 584 726 247 315 896 915 438 589 462 731 762

E3 12=2×6=3×4 12=3+3+3+3 3+4+3+4=14 2+3+2+3+2+3=15

E4 $a+b=b+a$ $\left(a+b\right)+c=a+\left(b+c\right)$ $a\left(b+c\right)=ab+ac$

E5 $(x+3)(x+4)=x^2+7x+12$ $(y+a)(y-a)=y^2-a^2$ $(x+3)(x-3)=x^2-9$

E6 $y=ax^2+bx+c$ $x1=\dfrac{-b+\surd(b^2-4ac)}{2a}$ $x2=\dfrac{-b-\surd(b^2-4ac)}{2a}$

E7 $y=x^2$ $\dfrac{dy}{dx}=2x$ $y=4x^3$ $\dfrac{dy}{dx}=12x^2$ $\int e^x dx=e^x+c$ $\int\dfrac{1}{x}\,dx=\ln x+c$

E8 $\dfrac{d(uv)}{dx}=v\dfrac{du}{dx}+u\dfrac{dv}{dx}$ $\int u\dfrac{dv}{dx}dx=uv-\int v\dfrac{du}{dx}dx$

Fig 4. The equation stimuli items

The test stimuli included four practice items and eight equation items, shown in Fig 4. Each practice item was a list of all the symbols in the response grid but in particular arrangements that aimed to support participants familiarization with the grid. The equation items, shown in Fig 4, were designed to span a range of sophistication in mathematical competence, from number sequences to equations and on to calculus formulas.

Procedure

The trials were run on the individual computers of the experimenters with a conventional mouse for input. The mouse settings of the computers were adjusted to suit the preferences of each participant. Bespoke software with millisecond temporal resolution and accuracy (SMouseLog), programmed within the Investigator's lab, was used to present the response grid and to log interactions.

After familiarization with the experimental setup, the centre-click interaction was explained and was briefly practiced. The stimulus items were individually presented printed on a card placed immediately to the left of the computer at the same height as the response grid on the display. The practice items were given in the order and the presentation order of the equation items (Fig 4) was randomised. For each trial the participants were told to begin as soon as the item was revealed and to centre-click copy as quickly and as accurately as possible, and simply to continue without pausing if they made a mistake.

Results

The student experimenters collated the data from their participants using a pre-prepared spreadsheet that extracted the identity of the symbols from the computer logs and calculated Q1 and Q3 for the stimulus, RTC, All, RTC+stim and stim-RTC pauses as defined above. The dataset comprised approximately 13,200 data points including symbols clicks and RTC actions, with each participant providing about 600 data points, depending on whether they used symbols 'd' and 'x' or the combined 'dx' for copying equation items E7 and E8.

The participants quickly became used to centre-click input. Inspection of the time series graphs of the pauses for the first practice item revealed that all participants' shortest pauses had asymptoted to a value typical of their individual shortest pause for later items, implying that centre-click input had rapidly become automatic.

Pause magnitudes

Fig 5 shows sample data of the $Q3_{All}$ pauses for the least and most competent participants, P1 and P22, respectively, on equation item E4. The pauses range from about 500 ms to over 5000 ms (above the graph). The magnitudes of both participants' shortest pauses are approximately equal but P1's curve has many more long pauses than P22. As this graph is typical of other participants, Q1 pauses should in general be similar among participants and Q3 should vary with competence and potentially be a good measure of competence. The '#' and '•' labels along the x-axis stand for RTC actions. The '•' labels also indicate where mathematically meaningful chunk boundary occur in the stimulus: specifically, breaks between equations or following the equal sign. P22's longest pauses largely coincide either with a • RTC label or the immediately following symbol, whereas as for P1 few of these boundaries are associated with long pauses and many long pauses occur elsewhere. This implies that P22's mathematical knowledge at the level of the equation structure substantially influenced P22's performance, whereas P1's profile is not consistent with the use of such knowledge. Although P22's longest pauses are associated with chunk boundaries, some of these pauses occur with symbol clicks and others with RTC actions, which suggests the investigation of different pauses measures is worthwhile.

Fig 6 shows the Q3 pause measure for each of the practice and equation test items. There is little variability in the click durations. For the first practice item, Pr1, which involved clicking symbols in the order of appearance the response grid, there is no difference between the All, symbol and RTC pauses, which implies participants are largely at-

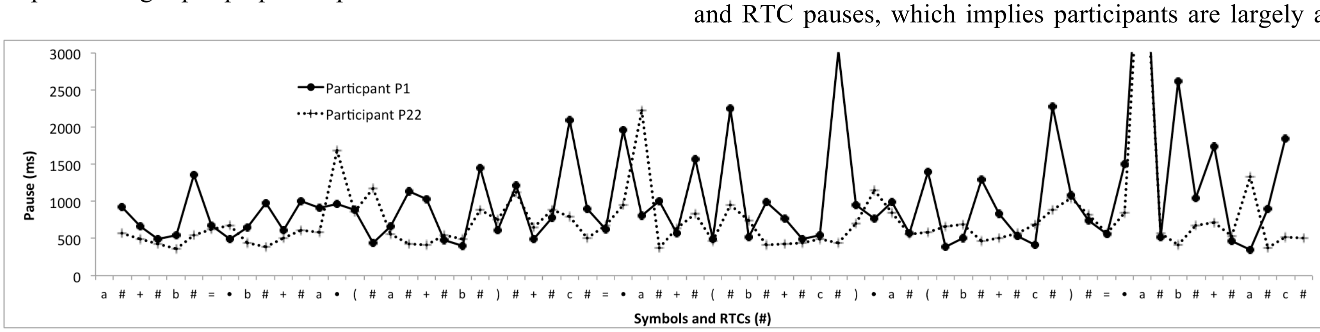

Fig 5. Pauses for symbol and RTC actions (#,•) by the least (P1) and most (P22) competent participants.

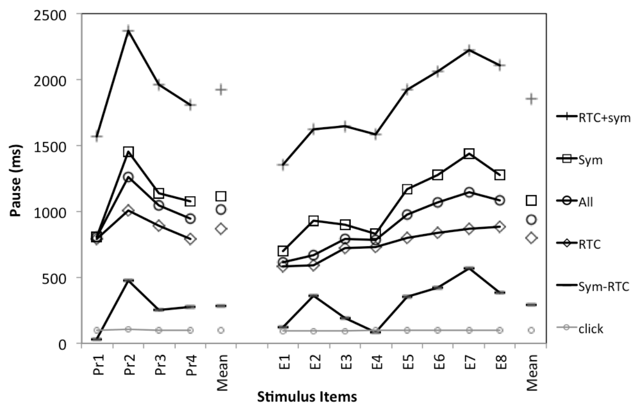

Fig 6 Various Q3 pauses across stimulus items

tending to grid location rather than symbol identity. Q3 pauses decreases substantially across practice items Pr2 to Pr4, which suggests participants became well familiarized with the symbols themselves. In the equation items a general trend exists for all measures to increase with increasing sophistication of the equations, which is monotonic for $Q3_{RTC}$ and nearly so for $Q3_{ALL}$.

Chunk processing with before symbols or RTCs?

Does perceptual and cognitive processing of chunks tend to occur before the symbol clicks or RTC actions? In Fig 6, across both practice and equation items RTC pauses are shorter than the symbol pauses and over the equation items the rate of increase is greater for $Q3_{sym}$ (range=742 ms) than for $Q3_{RTC}$ (range=301 ms). Across all 88 practice trials 72% of $Q3_{sym}$ pauses were greater than the $Q3_{RTC}$ pauses. In the 176 equation trials 75% of the $Q3_{sym}$ pauses were greater than the $Q3_{RTC}$ pauses. Aggregating over the trials for each participant the $Q3_{sym}$ pauses for participants were longer for symbols than $Q3_{RTC}$ pauses: practice items, mean $Q3_{sym}$ = 1142 (S.D. = 267) versus mean $Q3_{RTC}$ = 886 (S.D. = 201) ms; equation items mean $Q3_{sym}$ = 1072 (S.D. = 241) versus mean $Q3_{RTC}$ = 417 (S.D. = 85) ms, which by one-tail T tests (N=22) are both significant at p<.001. All this taken together implies that chunk processing is more associated with symbols than RTCs. Thus, it is expected that measures for RTC alone will not be good predictors of competence.

Competence correlations

Pearson's Correlations between the pause measures and competence scores of participants' were computed for each test item. In all cases N=22 (df=20) and the R values for standard critical values are: p=.05, R=0.360; p=.01, R=0.492. For the purpose of exposition, correlations at these levels will be called *weak* and *strong*.

Although six weak correlations for particular test items were found for first quartile measures, $Q1_{All}$, $Q1_{sym}$, $Q1_{RTC}$, $Q1_{RTC+sym}$, and $Q1_{sym-RTC}$, they are in a minority among the 60 measures, which is much as expected by chance. No strong correlations were present. For completeness correlations of competence test scores with the durations of mouse button presses were also calculated, $Q1_{click}$, $Q3_{click}$ and Q3-$Q1_{click}$. Give the 24 comparisons across the eight equations, the two weak correlations that were found can be attributed to chance. Reassuringly, there are no unexpected effects in participants' underlying performance.

Table 1 shows correlations values for various measures all against the participants' competence scores. The table includes correlations for the mean of the overall Q3 pauses found by aggregating over practice items or over equation

Table 1. Correlations of Q3 pause measures with competence.

	Pr1	Pr2	Pr3	Pr4	Pr mean	E1	E2	E3	E4	E5	E6	E7	E8	E mean
$Q3_{All}$	-0.14	-0.57 **	-0.50 **	-0.42 *	-0.50 **	0.29	0.29	0.05	-0.32	-0.43 *	-0.53 **	-0.26	-0.50 **	-0.31
$Q3_{sym}$	0.01	-0.68 **	-0.48 *	-0.34	-0.52 **	-0.07	-0.53 **	-0.33	-0.43 *	-0.38 *	-0.49 **	-0.27	-0.42 *	-0.48 *
$Q3_{RTC}$,	-0.18	0.06	-0.05	-0.18	-0.10	0.44 *	0.36 *	0.18	-0.10	-0.27	-0.13	-0.10	-0.08	0.01
$Q3_{RTC+sym}$,	-0.02	-0.54 **	-0.48	-0.27	-0.39 *	0.29	-0.17	-0.10	-0.23	-0.41 *	-0.46 *	-0.27	-0.37 *	-0.31
$Q3_{sym-RTC}$	0.09	-0.54 **	-0.33	-0.20	-0.35	-0.35	-0.51 **	-0.38 *	-0.29	-0.14	-0.31	-0.22	-0.27	-0.38 *

* – p<.05, 'weak' correlation, ** – p<.01 'strong' correlation

Table 2. Correlations of IQR pause measures with competence.

	Pr1	Pr2	Pr3	Pr4	Pr mean	E1	E2	E3	E4	E5	E6	E7	E8	E mean
IQR_{All}	-0.14	-0.48 *	-0.59 **	-0.60 **	-0.56 **	0.14	0.12	0.08	-0.47 *	-0.54 *	-0.62 **	-0.33	-0.57 **	-0.53 **
IQR_{sym}	0.05	-0.65 **	-0.50 **	-0.41 *	-0.53 **	-0.22	-0.61 **	-0.38 *	-0.48 *	-040 *	-0.49 *	-0.21	-0.39 *	-0.51 *

Table 3. Correlations for basic mathematical skills normalization measures, $Q3^N$, with competence.

	E3	E4	E5	E6	E7	E8	E mean
$Q3^N{}_{All}$	-0.23	-0.56 **	-0.62 **	-0.59 **	-0.39 *	-0.55 **	-0.60 **
$Q3^N{}_{sym}$	-0.06	-0.13	-0.15	-0.36	-0.08	-0.25	-0.21

items. None of the measures has a majority of strong correlations for the eight test items and none have strong correlation for their aggregate Q3 measures. In terms of numbers of strong and weak correlations $Q3_{sym}$ performs best followed by $Q3_{All}$. $Q3_{RTC}$ is the worst. $Q1_{RTC+sym}$, and $Q1_{sym-RTC}$ are in the middle, which is not surprising as they combine symbol and RTC pauses, although in different ways.

The IQR measure attempts basic-skill normalization on Q3 pauses by deducting Q1 pauses, on an item-by-item basis for each participant. IQR_{RTC}, $IQR_{RTC+sym}$, and $IQR_{sym-RTC}$ correlations with competence scores show little improvement over their respective Q3 measures. Table 2 gives the correlations of IQR_{All} and IQR_{sym}. IQR_{sym} normalization has relatively little impact on Q3sym. IQR_{ALL} normalization increases the strength of the correlations for items E4 to E8, compared to $Q3_{All}$ alone, so the IQR_{ALL} correlation for the mean of all the equations items is now strong.

Basic mathematical skills normalization, $Q3^N$, is computed by subtracting a measure of basic math skills from $Q3_{All}$, $Q3^B$ in Fig 3. Conveniently, the mean of Q3 pauses for equation items E1 and E2 provides such a measure, as these a numerical items. To be precise, this measure ($Q3^B$) is denoted by $Q3_{All,E1E2}$. For the normalization it is subtracted from each of the other equation items E3 to E8, individually (i.e., $Q3^N_{All,E3} = Q3_{All,E3}\text{-}Q3_{All,E1E2}$, $Q3^N_{All,E4} = Q3_{All,E4}\text{-}Q3_{All,E1E2}$, etc). The same was done for $Q3_{sym}$ for E3 to E8. Table 3 shows the correlations of these measures with competence score. For $Q3^N_{sym}$ no relation holds. For $Q3^N_{All}$, just one equation test item has no correlation (E3), one is weak (E7) and all the rest are strong (S4, S5, S6 and S8). The correlation for the average pauses across all test items is now $Q3^N_{All} =$ -0.6 compared to $Q3_{All}$ = -0.31.

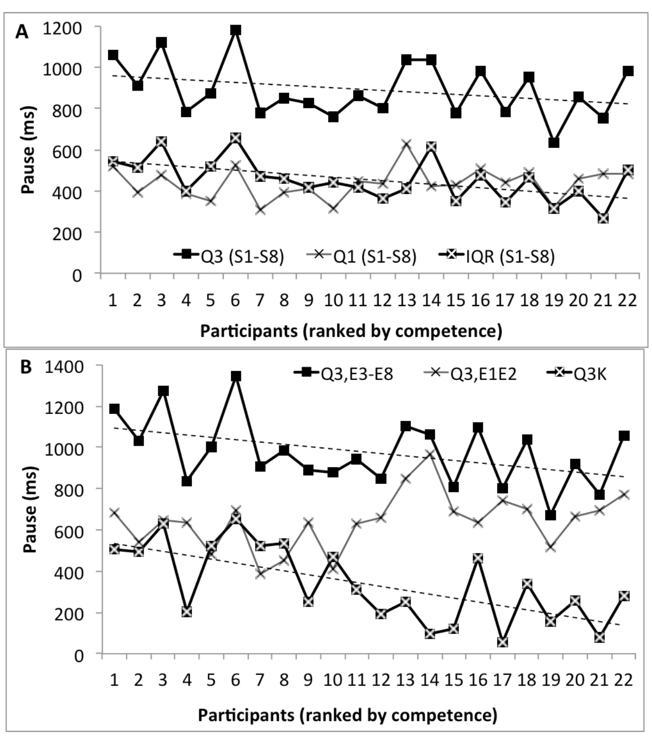

Fig 7. Normalizations of pause measures:
(A) IQR_{ALL} ($Q3\text{-}Q1_{ALL}$); (B) $Q3^N_{ALL}$, basic maths skills.

Fig 7 reveals why the basic mathematical skill normalization works better than the elementary skill normalization. Fig 7A shows the $Q3_{ALL}$ and $Q1_{ALL}$ for all equations items and the resulting normalized IQR_{ALL} curve. The dashed lines are least-squares best-fit lines for $Q3_{ALL}$ and IQR_{ALL}. The deviations of $Q1_{ALL}$ values from the best-fit lines shows there are substantial individual differences, but the overall shape of $Q1_{ALL}$ its similar to $Q3_{ALL}$, which means it is encoding largely similar individual differences to as those already captured by $Q3_{ALL,}$ so normalizing using $Q1_{ALL}$ to compute IQR_{ALL} does a little to improve upon $Q3_{All}$. In Fig 7B the shape of $Q3_{ALL,E1E2}$ (pauses for the two number items) is quite different to $Q3_{All,E3\text{-}E8}$ (pauses for equations proper), so the difference found by subtracting the former from the letter to give $Q3^N_{All}$, does adjust $Q3_{All,E3\text{-}E8}$ for individual differences that it does not already encode. In turn, this yields a stronger the correlation with competence score, as shown by the gradients of the best-fit lines.

Discussion

This experiment addressed questions about the use of the paused-based temporal chunk signal as a measure of mathematical competences in the simple task of copying equations. A reasonable strength correlation of -0.60 was found between mathematical competence and the third quartile pause measure after normalization for basic mathematical skills ($Q3^N$). For the normalized interquartile range measure (IQR_{All}) a correlation of -0.53 was obtained. These are weaker than the -0.72 correlation for a simple Q3 measure and the -0.73 correlation for an IQR measure obtained by Cheng (2014) for handwritten copying of equations. Apart for the intrinsic difference in the modes of interaction, a possible factor in the weaker correlations found here is a less uniform distribution of participant' competence than in the Cheng (2014) sample; here participants with lower mathematical competence were somewhat over represented. Nevertheless, the strength of the correlations (especially $Q3^N_{All}$) is sufficient to claim that the centre-click interaction exhibits the temporal chunk signal with sufficient strength that it can be used to measure mathematical competence. In turn, this demonstrates that temporal chunk signal extends beyond normal freehand writing and drawing to use with conventional mouse operated computer interfaces. In contrast to previous studies that used typewriting, centre-click copying does not require aggregation over large amounts of data to differentiate levels of competence (cf., the studies cited by Salthouse, 1986), nor it is hampered by large individual differences in typing skills that mask the signal when assessments are limited to a small number of test items per participant (c.f., Ismail, 2010).

The specific design of the response grid (Fig. 1) appears to be important to the efficacy of centre-click copying used to measure mathematical competence. In an earlier pilot experiment, in which no strong correlations to mathematical competence were found, the distances of the symbols to the center in the response grid was far less uniform than in the present experiment. However, the distances to the symbols are not equal in Fig 1, so it will be interesting to investigate whether this is still a substantial source of noise for the Q3 measures.

Following Cheng (2014), Cheng & Rojas-Anaya (2007), Zulkifli (2013) and van Genuchten & Cheng (2009), it has again been shown that competence in a knowledge-rich cognitive tasks can be measured using a simple stimuli copying activity. The more knowledgeable participants exhibit shorter pauses (Figs 5 & & 7), because they possess relevant chunks that likely support their more rapid encoding of larger (meaningful) patterns from the stimuli. The claim that copying has some potential as a general method for measuring competence is further supported by the success of centre-click interaction. As the predicted order of difficulty of the equation test items appears to be reflected in the aggregate measures of pauses, Fig 6, this adds weight to the claim that the process of copying and the Q3 pauses therein are related to participants level of understanding of each target stimulus.

Centre-click copying involves the RTC, return to center, actions and symbol clicks, so the time associated with the processing of chunks might be associated with symbol and RTC pauses in different ways. Five theoretically possible pauses measures were defined in the introduction: $Q3_{ALL}$, $Q3_{sym}$, $Q3_{RTC}$, $Q3_{RTC+sym}$, $Q3_{sym\text{-}RTC}$. The values of $Q3_{sym}$ are substantially longer than $Q3_{RTC}$ pauses, Fig 6, which suggests that $Q3_{sym}$ ought to have had the strongest correlations with the competence test scores. However, just the pooled pauses including symbol clicks and RTC actions gave strong correlations (Table 2). Of the other measures, there was a hint that $Q3_{sym}$ might reflect the temporal chunk signal, but no other meaningful patterns of correlations were observed. All this suggests that there are substantial differences in the way individual participants distribute the processing of chunks within centre-click copying, despite the relatively simple nature of the underlying task.

In Cheng (2014) the normalization for elementary skills and basic mathematical skills did not not improve the Q3 pause measures. Here, in contrast, the IQR_{All} correlations were significant and did improve the measure and even more so with the normalization for basic mathematical skills, $Q3^{N}_{All}$. This occurred for the overall aggregate measure but more impressively it was seen at the level each equation test item (Table 3 cf. Table & 2). The improvement for $Q3^{N}_{All}$ worked here because the distribution of $Q3_{All,E1E2}$ pauses across participants for number test items reflects individual differences not seen in the $Q3_{All,E3E8}$ distribution for equations proper (Fig 7). In Cheng (2014) the equivalent number and equation test item distributions were similar, which implies that it was a less sensitive measure of underlying mathematical skills. The overall implication, of this and the point in the previous paragraph, is we must understand to what extent, and how, individual differences are manifested in different techniques to measure the temporal chunk signal when evaluating competence.

Acknowledgments

My thanks go to Shona McLeod, Maya Stanford, Matthew Bodell and Mikael Fodor their conscientious and thorough collection and collation of the data.

References

Cheng, P. C-H. (2014). Copying equations to assess mathematical competence: An evaluation of pause measures using graphical protocol analysis. In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds.), *Proc. of the 36th Annual Meeting of the Cognitive Science Society* (pp. 319-324). Austin, TX: Cog. Sci. Soc.

Cheng, P. C-H., & Rojas-Anaya, H. (2007). Measuring Mathematical Formula Writing Competence: An Application of Graphical Protocol Analysis. In D. S. McNamara & J. G. Trafton (Eds.), *Proc. of the 29th Annual Conf. of the Cognitive Science Society* (pp. 869-874). Austin, TX: Cog. Sci. Soc.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Science*, 24(1), 87-114.

Ismail, H.B. (2010). Using pauses in typing to measure language competence. Unpublished MSc dissertation, Department of Informatics, University of Sussex.

Obaidellah, U. H., & Cheng, Peter C-H. (2009). Graphical production of complex abstract diagrams: drawing out chunks and schemas. In N. Taatgen & H. v. Rijn (Eds.), *Proc. of the 31st Ann. Conf. of the Cognitive Science Society*. Austin, TX: Cog. Sci. Soc.

Roller, R., & Cheng, P. C-H. (2014). Observed strategies in the freehand drawing of complex hierarchical diagrams. In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds.), *Proc. of the 36th Annual Meeting of the Cognitive Science Society* (pp. 2020-2025). Austin, TX: Cog. Sci. Soc.

Salthouse, T.A. (1986). Perceptual, Cognitive, and Motoric Aspects of Transcription Typing. *Psychological Bulletin*, 99(3), 303-319.

van Genuchten, Erlijn, & Cheng, Peter C. H. (2009). Missing working memory deficit in dyslexia: children writing from memory. In N. Taatgen & H. v. Rijn (Eds.), *Proc. of the 31st Annual Conf. of the Cognitive Science Society* (pp. 1674-1679). Austin, TX: Cog. Sci. Soc.

van Genuchten, E., & Cheng, P. C.-H. (2010). Temporal chunk signal reflecting five hierarchical levels in writing sentences. In S. Ohlsson & R. Catrambone (Eds.), *Proc. of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 1922-1927). Austin, TX: Cognitive Science Society.

Zulfliki, P. A. M. (2013). *Applying pause analysis to explore cognitive processes in the copying of sentences by second language users.* Unpublished PhD thesis, Department of Informatics, University of Sussex.