Calculation Fluency Test

Carla Sowinski, Kristina Dunbar, and Jo-Anne LeFevre
Technical Report: Math Cognition Laboratory
Carleton University

BACKGROUND

Test Development

The Calculation Fluency Test (CFT) was developed in the Math Cognition Lab at Carleton University. It was modeled on two sub-tests of the French Kit of Cognitive Factors (French, Ekstrom, & Price, 1976), a set of various measures of mental processing. The Addition Sub-test comprises rows of three-term addition problems (arranged vertically) whereas the Subtraction/Multiplication sub-test has alternating rows of double-digit subtraction problems and double times single-digit multiplication problems.

Although we used these two French Kit subtests for many years to index **arithmetic fluency** (e.g., Smith-Chant & LeFevre, 2002; LeFevre et al., 2006), that is, the speed and accuracy of multi-digit calculation, increasingly we found that participants did poorly. On the subtraction-multiplication subtest they often did not finish a single subtraction row, and thus were not evaluated on multiplication. In general, performance had been declining fairly steadily from the early 1990s (LeFevre et al., 2014). Accordingly, we develop a simplified fluency test that allows us to contrast performance across the three operations, and to avoid the confound of having subtraction and multiplication combined on one page. Also, in the revised version, addition was limited to two-term problems (rather than three terms) so that students would have the chance to try more problems. The administration was also simplified: We reduced the number of pages from four (two in each subtest) to three (one for each operation), and reduced the time per page to one minute from two. Thus, total administration time was reduced from 10 minutes to five minutes, allowing time for other measures to be used as part of the general skills assessment, as required. The goal of the present report is to provide information about the inter-correlations among the three subtests in the revised test and to determine whether certain patterns that were observed with the original measure would also be found with the revised test. A large group of participants completed the measure over several years; most also provided basic demographic and educational information.

What does the CFT measure?

The test is designed to assess the fluency with which, for each of addition, subtraction, and multiplication, participants (a) access and implement the relevant calculation procedures and (b) retrieve or compute the answers to the component arithmetic facts. Participants are not told what solution procedures to use, on the assumption that they already have a familiar procedure available. The practice component does allow them to ask questions, however, to clarify the instructions or refresh their memories about the calculation process. Problems are all presented in vertical orientation, which may prime a standard "right-to-left" solution algorithm, at least for participants educated in North America. Participants are also told that they can use the extra space on the paper for "rough work", so they are not forced to rely solely on mental arithmetic.

Consider the problem 29 + 43. In the standard algorithm, a solution might involve adding 9 + 3, writing '2' in the right column of the answer box, carrying the "10" (possibly writing a "1" in the left column), adding 2 + 4 + 1 and writing 7 in the left column. Alternatively, another individual might round 29 to "30", add 43 to get 73, then subtract 1 and write down 72. No measure of how the individuals solve the problems is collected.

METHOD

Description of Test

This test comprises three pages of two-term arithmetic problems. The first page has double-digit addition problems (e.g., 52 + 19), the second page has double-digit subtraction problems (e.g., 84 – 47), and the third page has double times single-digit multiplication problems (e.g., 67 x 4). The test is shown in Appendix 1. A PDF version can be downloaded from: www.carleton.ca/cacr/mathlab/measures.

Procedure

Participants are given one minute per page. They are told to solve the problems quickly and accurately. Each page has six rows of problems with 10 problems per row. The test is scored as the total number of items solved correctly across the three pages. Participants solve six practice problems of each type before starting the test. Thus, total administration time is approximately five minutes.

Sample

Participants in nine different experiments (N = 441) completed the CFT and the MBIQ as part of the requirements for those experiments. Age (in years) was available for 335 of the participants. As shown in Table 1, most participants were between 18 and 22 years of age, consistent with the demographic of the population of university students.

Many individuals also indicated the country in which they had received their elementary education. We assumed, based on considerable cross-cultural research, that much of an individual's arithmetic skill is developed during the first six grades of elementary school. Because of the demographics of the undergraduate population, we categorized three groups: Those who received their elementary education in Canada, those who received their education in China, and a third group that comprised all other countries. Participants in the "other" group were from a wide-range of countries including: Iran, United Arab Emirates, Nigeria and Saudi Arabia. In our lab, we often compare groups of Chinese-educated individuals to groups of Canadian-educated individuals, which is why the Chinese-educated group is relatively large. We also know from much of our previous work that this group will have superior arithmetic skills, as measured by the original French kit sub-tests (e.g., LeFevre & Liu, 1997; Xu, Wells, Imbo, & LeFevre, 2014). Thus, we predicted that the Chinese-educated group would have higher scores.

In all groups, the small number of older students resulted in a positive skew, with the smallest effect on the Chinese group. As noted, the range of ages in the Chinese group was much more restricted because there would be very few Chinese-educated students who were outside the typical undergraduate age range.

Table 1. Descriptive Statistics for Age for the Complete Sample and Three Subsamples. Group refers to the country in which participants received their elementary education.

Group (N)	*N*	*% female*	*Min*	*Max*	*Mdn*	*M*	*SD*	*Skew*	*Kurtosis*
All	335	57	17	60	20	21.74	5.70	3.85	18.92
Canadian	199	58	17	60	19	20.92	5.75	4.29	21.41
Chinese	100	68	18	31	22	22.74	3.16	.66	-.44
Other	81	45	17	59	20	22.54	7.49	3.32	12.15

Results

Correlations. The correlations between age and scores on each of the subtests of the CFT are shown in Table 2 for the sample of participants who provided age information. Age is correlated with scores on each subtest, because older individuals had higher scores. LeFevre et al. (2014) found a similar pattern for the original French Kit test. However, only a very small number of participants were older than 30 so the correlation may be attenuated by the non-normality of the age distribution. Further, participants who were older than 22 represent a mixture of undergraduate and graduate students. Age is only one of many possible influences on the distribution of performance.

Table 2. Correlations for Age and Calculation Fluency Test (CFT) pages

	1	2	3	4	5
1. Age	-				
2. CFT Addition	.21**	-			
3. CFT Subtraction	.20**	.84**	-		
4. CFT Multiplication	.20**	.79**	.78**	-	
5. CFT Total	.22**	.95**	.94**	.92**	-

** Correlation is significant at the .01 level (2-tailed)

Total score reliability. As shown in Table 2, the subtests were highly inter-correlated with one another. Accordingly, a total score on the CFT was created by summing the number correct for addition, subtraction, and multiplication. The descriptive statistics for this total score are shown in Table 3. The internal reliability of this summed scored is high, with Cronbach's alpha of .924 (n = 440). No substantial skew was observed, and medians were similar to means for all groups, indicating that the measure is approximately normally distributed. Note that the mean score for the Chinese participants was approximately twice that of the Canadian participants.

Table 3. Descriptive Statistics for Total CFT Scores for the Complete Sample and Three Subsamples

Group	*N*	*Min*	*Max*	*Mdn*	*M*	*SD*	*Skew*	*Kurtosis*
All	441	3	107	36	39.03	18.70	.81	.57
Canadian	242	3	84	30	31.62	13.53	.63	.46
Chinese	103	30	107	58	58.98	16.72	.53	.09
Other	95	13	91	33	36.36	16.58	1.09	.49

As shown in Tables 4 through 6, performance varied across pages, with participants generally scoring higher on addition than subtraction (M = 17 vs. 13), both of which were higher than multiplication (M = 8). The range of scores was similar across operations. Some participants had no correct answers, and thus minimum scores of 0 were observed for subtraction and multiplication. Correlations among the scores were high, however, as shown in Table 2. Anecdotal reports from the experimenters who collected the data suggest that some participants either did not know, or had forgotten, how to solve vertically presented multi-digit arithmetic problems. None of the scores showed much skew, and medians were similar to means, indicating that these individual scores are approximately normally distributed.

Table 4. Descriptive Statistics for CFT Addition Scores for the Complete Sample and Three Subsamples

Group	*N*	*Min*	*Max*	*Mdn*	*M*	*SD*	*Skew*	*Kurtosis*
All	441	2	43	16	17.28	7.4	.71	.27
Canadian	243	2	37	14	14.73	5.9	.65	.61
Chinese	103	15	43	23	24.32	6.58	.56	-.17
Other	95	5	38	14	16.15	6.87	1.07	.49

Table 5. Descriptive Statistics for CFT Subtraction Scores for the Complete Sample and Three Subsamples

Group	*N*	*Min*	*Max*	*Mdn*	*M*	*SD*	*Skew*	*Kurtosis*
All	440	0	39	12	13.16	6.34	.79	.85
Canadian	243	0	24	10	10.85	4.69	.21	-.16
Chinese	103	5	39	18	19.23	6.43	.43	.27
Other	95	4	32	12	12.49	5.65	1.04	1.24

Table 6. Descriptive Statistics for CFT Multiplication Scores for the Complete Sample and Three Subsamples

Group	*N*	*Min*	*Max*	*Mdn*	*M*	*SD*	*Skew*	*Kurtosis*
All	440	0	33	8	8.54	6.28	.82	.33
Canadian	242	0	24	5	5.94	4.55	.93	.63
Chinese	103	3	33	15	15.43	5.38	.53	.61
Other	95	0	27	7	7.72	5.47	1.03	.49

Educational background. Total scores on the CFT were analyzed in a 3(Group: Canadian, Chinese, Other) ANOVA. Scores varied with group, $F(2, 391) = 68.90$, $p < .001$. Post-hoc tests (Bonferroni correction) showed that Canadian participants had lower scores than those educated in other countries (ps $< .05$), and Chinese-educated participants had higher scores than the other two groups (ps $< .001$). All three groups solved the most equations correctly on addition, followed by subtraction, and then multiplication.

Sex. There was no difference between women and men on the total score, $F(1, 433) = .163$, $p > .05$.

Construct validity. To determine whether the CFT scores showed expected relations with self-report measures, total CFT scores were correlated with participants' responses to several questions from the Math Background and Anxiety Questionnaire (see website). For each question, participants indicate how nervous or how likely they are to avoid the activity, where 1 is extremely nervous/avoidant and 7 is not at all nervous/avoidant. As expected, total CFT scores were correlated with self-reported math anxiety, $r(391) = ..22$, $p < .01$, meaning that participants reporting less anxiety had higher scores on the CFT. Similarly, there was a correlation between math avoidance behaviour and CFT scores, $r(392) = .28$, $p < .01$, such that lower levels of self-reported math avoidance correlated with higher CFT scores. However, participants' reports of their feelings about reading avoidance and scores on the CFT were not

significantly correlated, $r(360) = -.09$ and although the correlation between the CFT score and avoidance of writing was significant, it was low, $r(358) = -.15$, $p < .01$.

DISCUSSION

The goals of the reformulation of the original fluency test were met, and there is evidence that the new version shows the same patterns as observed in the original. The correlations among the pages of the test are high, such that the internal validity is excellent for the total score. Chinese-educated participants scored much higher than Canadian-educated participants on all subtests and on the overall score. Participants educated in other countries scored significantly higher than Canadian-educated participants. Performance was correlated with participants' age, although the effect size was modest. Finally, performance correlated with math avoidance and anxiety measures at a similar level as in previous research using the original measure (LeFevre et al., 2014). All of these patterns suggest that the CFT captures a similar index of arithmetic fluency as the original fluency sub-tests. Finally, in several experiments the CFT has shown good concurrent validity with other measures of arithmetic skill (Bourassa, 2014; Sowinski & LeFevre, 2014). Thus, we feel confident that the revised fluency measure is a reliable and valid index of adults' multi-digit arithmetic skill. Note that the measure was designed for use in experiments where a quick measure of fluency would be useful. We do not have any information, at this point, on whether the measure would be useful beyond an experimental context.

References

LeFevre, J.-A., Penner-Wilger, M., Pyke, A. A., Shanahan, T., & Deslauriers, W. A. (2014). *Putting two and two together: Declines in arithmetic fluency among young Canadian adults, 1993 to 2005*. Technical Report: Math Cognition Laboratory. Ottawa, ON. Retrieved from http://www.carleton.ca/ics/research/technical-reports/view-reports/