
Excluded Letter Fluency Test (ELF): Norms and Test-Retest Reliability Data for Healthy Young Adults

E. Arthur Shores

Macquarie University, Australia

Jane R. Carstairs

University of Wolverhampton, United Kingdom

John R. Crawford

University of Aberdeen, United Kingdom

Normative and reliability data for the Excluded Letter Fluency (ELF) Test are provided. A stratified random sample of 399 healthy young adults aged 18 to 34 years from Sydney, Australia, completed the ELF Test as well as a full-length WAIS-R, as part of a larger battery of tests. After a 1-year interval 99 of these individuals were retested on the same forms of the tests. The influence of age, sex and education was investigated on the ELF and only education was found to have a significant overall effect on the total scores. However, gender was found to have an effect on the error scores, with males making more rule-breaks than females. Tables are provided for converting ELF raw scores, corrected for years of education, to standard scores with 90% and 95% confidence intervals for both test and retest purposes. A table for calculating the base rate of errors, for males and for females, on the ELF is also provided.

Normative data sets exist on numerous forms of verbal fluency tests (Lezak, Howieson, & Loring, 2004). The concept of measuring verbal fluency can be traced back to the Thurstone Word Fluency Test, which required a written response (Thurstone, 1938) and an oral form, which has come to be known as the FAS Test or Controlled Oral Word Association Test (Bechtoldt, Benton, & Fogel, 1962; Benton, 1967; Spreen & Benton, 1969). Most forms of the verbal fluency task have required patients to provide, in a given period of time, as many words as they can think of beginning with a particular letter (initial-letter fluency) or belonging to a particular category such as animals or fruits (semantic fluency).

The popularity of initial-letter fluency tests in clinical neuropsychology may be accounted for by the sensitivity of such tests to linguistic impairment,

particularly following dominant frontal system or executive dysfunction (e.g., Elfgrén, Ryding, & Passant, 1996). It has been argued that semantic fluency tasks are relatively insensitive to executive dysfunction because everyday word retrieval is conducted on the basis of semantics and this is therefore a relatively non-effortful task. In contrast, initial-letter fluency requires a novel retrieval strategy and imposes a heavier monitoring demand to prevent production of illegitimate, semantically related words. Recently, however, a more demanding form of verbal fluency, excluded letter fluency, has been demonstrated to be sensitive to cognitive impairment in patients with traumatic brain injuries (Crawford, Wright, & Bate, 1995). The excluded letter fluency test requires patients to generate as many words as they can that do not contain certain letters. In the Crawford et al. study 24 patients and

Address for correspondence: E. Arthur Shores, Department of Psychology, Macquarie University, Sydney NSW 2109, Australia. E-mail: ashores@psy.mq.edu.au

24 matched control participants were required to generate words that did not contain the letters A, then E, then I. This task therefore is further removed from routine retrieval and places heavier demands on the monitoring process. One of the hypotheses tested in the Crawford et al. study was that production scores on the different fluency tasks would differ significantly for the traumatic brain injury (TBI) group compared with the healthy controls. Although Crawford et al. found a significant main effect for group (TBI and healthy controls) and task (semantic fluency, initial-letter fluency and excluded letter fluency) there was no significant interaction. This indicates that the different fluency tasks were not differentially sensitive to TBI. However, the excluded letter fluency test provoked significantly more error scores in the TBI group than controls, whereas the other fluency tasks showed no significant differences in error scores by group. This suggests that the excluded letter fluency test has potential as a sensitive measure of self-monitoring, a salient feature of everyday functioning following head injury, not usually easily elicited by formal tests currently in use (Ponsford & Kinsella, 1992).

Miller (1994), on the basis of finding a high correlation between an index of verbal intelligence and verbal fluency ($r = .87$), provided a regression equation to predict initial-letter fluency from IQ. However, the normal group used was small ($N = 30$) and only three verbal subtests of the WAIS were used to estimate verbal IQ. Crawford, Moore and Cameron (1992), using a larger sample of participants ($N = 142$) free of neurological or psychiatric disorder, provided a regression equation to predict an individual's expected ('premorbid') initial-letter fluency from performance on the National Adult Reading Test (NART). In their study a relatively high correlation between an estimate of IQ (NART errors) and verbal fluency was also found ($r = .67$).

Normative data with demographic corrections have been provided by Ruff, Light, Parker, and Levin (1996) for parallel forms of an initial verbal fluency task. They used the initial letters CFL and PRW as the alternate forms. Coefficient alpha was reported as $r = .83$ and test-retest reliability after 6 months was reported to be $r = .74$. They did not report correlations between IQ and fluency and did not provide information in relation to the standard errors of measurement.

No norms have previously been reported for an excluded letter fluency test. Furthermore, no studies appear to have investigated the relationship between an excluded letter fluency task and intelligence. If it were found that suitably high correlations existed

between these measures, clinicians could then use IQ to predict excluded letter fluency performances.

The present study reports normative data for an excluded letter fluency test derived from a relatively large group of healthy young adults, who also had been administered a full-length WAIS-R. These were participants in the Macquarie University Neuropsychological Normative Study (MUNNS; Carstairs & Shores, 2000). Test-retest reliability data also are reported from 99 of these participants who were retested after a period of one year. Correlations between IQ and excluded letter fluency test performance are reported as well as correlations between excluded letter fluency test performance and the Cambridge Contextual Reading Test and the Australian Contextual NART Reading Test. Standard errors of measurement for test and retest purposes are provided as are base rates of error scores.

Method

Participants

The MUNNS normative sample comprised 399 randomly selected healthy adults (193 males, 206 females) aged between 18 and 34 years (mean age 25.6 years, $SD = 5.0$ years), living in the Sydney metropolitan area. Sampling procedures were designed to provide a representative sample of healthy young Sydney adults within the target age range. The sampling strategy was developed in conjunction with the Australian Bureau of Statistics to be representative of 1991 census data in terms of sex, age, non-English speaking background, socioeconomic level, and educational background for the age range of 18 to 34 years. The sample frame was the Sydney metropolitan area which, for the purpose of the study, was defined as the geographical area within the 02 telephone prefix boundary as it existed in 1996. This area corresponds to six statistical subdivisions, with each subdivision having a consistent socioeconomic profile. Potential participants were contacted by telephone, using numbers derived from 'Oz on Disc', a CD version of the Sydney telephone directory. A random selection of telephone numbers was drawn from all numbers corresponding to each statistical subdivision.

Retest Sample. One hundred participants from the original sample were retested approximately 1 year following their initial testing. The retest sample was not required to meet all of the demographic requirements of the normative sample, but quotas were filled to ensure an equal number of males and females and representativeness with respect to age and language background.

Measures

The following tests were given as part of the larger MUNNS neuropsychological test battery.

WAIS-R (Wechsler, 1981). The full WAIS-R was administered and scored using the procedures in Wechsler (1981) and de Lemos (1981).

ELF (Crawford, Wright, & Bate, 1995). Instructions for administering and scoring the ELF are provided in Appendix A.

Cambridge Contextual Reading Test (CCRT; Beardsall & Huppert, 1994). Beardsall and Huppert investigated the extent to which putting NART words into sentences facilitated correct pronunciation. They found that both normal and demented subjects substantially improved their performance when the words were seen in context. In the MUNNS the CCRT was replaced with the Australian Contextual NART Reading Test part way through the study as this latter test was considered to be more appropriate in an Australian context. One hundred and thirty-five participants were tested on the CCRT.

Australian Contextual NART Reading Test (ACNRT). A contextual version of the Australian NART (AUSNART; Hennessy & Mackenzie, 1995) was developed for use in the MUNNS study. The original AUSNART words were put into context in the form of simple, meaningful sentences. These sentences are available in the Appendix to the article by Lucas, Carstairs and Shores (2003). Two hundred and forty-four participants were assessed on the ACNRT.

Results

Psychometric Properties of ELF

The internal consistency of the ELF (coefficient alpha), calculated from the total scores of the three trials: 'not A', 'not E', 'not I', was .84 ($N = 399$). The 1-year retest reliability was .67 ($N = 99$). A paired sample t test showed that there was no significant difference change in the retest score after one year ($t = .279$, $df = 98$, $p = .781$).

Correlations between the WAIS-R Full Scale IQ, Verbal IQ and Performance IQ and the ELF were .45, .44 and .35 respectively. The weak to moderate correlations between the WAIS-R and the ELF were considered insufficient to allow the meaningful or reliable prediction of excluded letter task fluency performance based on IQ.

Correlations between the total ELF raw score and the Cambridge Contextual Reading Test and the Australian Contextual NART were $-.29$ and $-.39$ respectively.

Derivation of Demographically Adjusted Norms

To investigate the influence of demographic variables on the raw scores of the ELF a three-way ANOVA was performed with age, sex and education as the independent variables. Participants were divided into two age groups (18 to 25, $n = 201$ and 26 to 34, $n = 198$). Three levels of education were used, less than 12 years ($n = 91$), 12 years ($n = 91$) and greater than 12 years ($n = 217$). A significant overall effect was found only for education ($F = 11.18$, $df = 2$, $p < .001$; eta squared = .055). Following the method used by Ruff et al., correction factors which adjust the means for each education group to a value equivalent to the overall mean were calculated. The means and standard deviations of the raw scores obtained on the ELF by education group, as well as the correction factors are provided in Table 1.

Calculation of Standard Scores, T Scores and Standard Errors

Raw scores were corrected for education. To ensure that the standard scores and T scores were normally distributed, corrected raw scores were first converted to a percentile rank based upon the cumulative frequency distribution. Then each percentile rank was transformed to its corresponding z score from the standard normal distribution. Finally, the z score distribution was converted to distributions having a mean of 100 and standard deviation of 15 (ELF standard scores) or a mean of 50 and standard deviation of 10 (ELF T scores). Standard scores, T scores and percentile ranks for corrected ELF raw scores are provided in Table 2.

For the ELF standard scores, standard errors of estimate and standard errors of prediction were calculated using Dudek's (1979) formulae 2 and 3. Note that we used coefficient alpha in deriving these values. Formula 2 calculates the standard error of estimate, which is used to derive the confidence interval for testing participants who are not part of the standardisation sample. Formula 3 calculates the standard error of prediction, which is used to derive the confidence intervals when participants are

TABLE 1
Mean Raw Scores on ELF by Education

Education	Mean	SD	Correction factor
Less than 12 years	42.6	13.7	+6
12 years	48.8	14.9	0
More than 12 years	51.5	13.7	-3

TABLE 2

ELF Raw Score to Standard Scores Conversions With 90% and 95% Confidence Intervals Around Estimated True Score Values for Test and Retest Purposes

Corrected Raw Scores	ELF Standard Scores	Confidence Intervals								T Score	%ile
		Test				Retest					
		90%		95%		90%		95%			
LL	UL	LL	UL	LL	UL	LL	UL				
	144+	—	—	—	—	—	—	—	—		
96	144	129	145	127	147	124	150	121	153	80	100
94- <u>95</u>	140	126	142	124	144	121	147	118	150	76	100
88- <u>93</u>	137	123	139	121	141	118	144	115	147	75	100
82- <u>87</u>	135	121	137	119	139	116	142	113	145	74	99
81	134	121	137	119	139	116	142	113	145	73	99
80	132	119	135	117	137	114	140	111	143	71	99
79	130	117	133	115	135	112	138	109	141	70	98
78	130	117	133	115	135	112	138	109	141	70	98
77	128	116	132	114	134	111	137	108	140	69	98
<u>76</u>	<u>127</u>	<u>115</u>	<u>131</u>	<u>113</u>	<u>133</u>	<u>110</u>	<u>136</u>	<u>107</u>	<u>139</u>	<u>68</u>	<u>97</u>
75	126	114	130	112	132	109	135	106	138	67	96
74	125	113	129	111	131	108	134	105	137	67	96
73	124	112	128	110	130	107	133	104	136	66	95
72	123	111	127	109	129	106	132	103	135	65	94
71	122	110	126	108	128	105	131	102	134	65	94
70	121	110	126	108	128	105	131	102	134	64	93
69	121	110	126	108	128	105	131	102	134	64	92
68	120	109	125	107	127	104	130	101	133	63	92
67	119	108	124	106	126	103	129	100	132	62	91
66	117	106	122	104	124	101	127	98	130	62	88
65	117	106	122	104	124	101	127	98	130	61	87
64	116	105	121	103	123	100	126	97	129	61	87
63	116	105	121	103	123	100	126	97	129	60	85
62	115	105	121	103	123	100	126	97	129	60	85
61	114	104	120	102	122	99	125	96	128	59	84
60	113	113	119	101	121	98	124	95	127	59	81
59	112	102	118	100	120	97	123	94	126	58	80
58	111	101	117	99	119	96	122	93	125	57	78
57	109	100	116	98	118	95	121	92	124	56	75
56	108	99	115	97	117	94	120	91	123	55	71
55	108	99	115	97	117	94	120	91	123	55	70
54	107	98	114	96	116	93	119	90	122	54	68
53	105	96	112	94	114	91	117	88	120	54	66
52	104	95	111	93	113	90	116	87	119	53	62
51	103	95	111	93	113	90	116	87	119	52	60
50	102	94	110	92	112	89	115	86	118	52	58
49	101	93	109	91	111	88	114	85	117	51	55
48	100	92	108	90	110	87	113	84	116	50	52
47	100	92	108	90	110	87	113	84	116	50	50
46	99	91	107	89	109	86	112	83	115	49	48
45	98	90	106	88	108	85	111	82	114	48	46
44	97	89	105	87	107	84	110	81	113	48	42
43	96	89	105	87	107	84	110	81	113	47	40
42	94	87	103	85	105	82	108	79	111	46	37
41	93	86	102	84	104	81	107	78	110	45	33
40	92	85	101	83	103	80	106	77	109	44	30
39	91	84	100	82	102	79	105	76	108	44	28

Table 2 (continued)

Raw Scores	Corrected ELF Standard Scores	Confidence Intervals								T Score	%ile
		Test				Retest					
		90%		95%		90%		95%			
		LL	UL	LL	UL	LL	UL	LL	UL		
38	89	83	99	81	101	78	104	75	107	43	25
37	88	82	98	80	100	77	103	74	106	42	22
36	86	80	96	78	98	75	101	72	104	41	20
35	85	79	95	77	97	74	100	71	103	40	17
34	84	79	95	77	97	74	100	71	103	39	15
33	82	77	93	75	95	72	98	69	101	38	13
32	81	76	92	74	94	71	97	68	100	37	11
31	79	74	90	72	92	69	95	66	98	36	9
30	78	74	90	72	92	69	95	66	98	35	7
29	77	73	89	71	91	68	94	65	97	34	6
28	76	72	88	70	90	67	93	64	96	34	6
27	75	71	87	69	89	66	92	63	95	33	5
26	74	70	86	68	88	65	91	62	94	33	5
25	72	68	84	66	86	63	89	60	92	31	4
24	70	67	83	65	85	62	88	59	91	30	3
23	69	66	82	64	84	61	87	58	90	29	2
22	68	65	81	63	83	60	86	57	89	28	2
21	65	63	79	61	81	58	84	55	87	27	1
<u>18-20</u>	63	61	77	59	79	56	82	53	85	25	< 1
<u>14-17</u>	60	58	74	56	76	53	79	50	82	24	< 1
<u>11-13</u>	56	55	71	53	73	50	76	47	79	20	< 1

Note: ELF = Excluded Letter Fluency Test, LL = Lower Limit, UL = Upper Limit. Underlined numbers are interpolated.

retested. Values are provided for the 68%, 90%, 95% and 99% confidence intervals in Table 3. Atkinson (1991) has recommended that when these values are used to calculate confidence intervals they should be placed around an individual's predicted true score, which is based on obtained results but regressed to the mean, rather than on the individual's obtained score. The following procedure may be used to convert an obtained score to a predicted true score: (i) multiply the obtained score by the reliability of the test (in this case .84); (ii) multiply the mean of the test (in this case 100) by 1 minus the reliability; and (iii) add the result of (i) and (ii).

For ELF standard scores, Table 2 provides 90% and 95% confidence intervals for test and retest, using the appropriate values from Table 3. It should be noted that as these values are based around the predicted true score, the obtained score does not necessarily correspond to the midpoint of the confidence interval.

ELF Error Scores

Error scores consisted of rule-breaks and perseverations totalled across the three trials of the test.

Spelling mistakes contributing to rule-breaks were not counted as errors. As with ELF total scores, the impact of demographic variables on ELF error scores was evaluated using a three-way ANOVA with age, sex and education as the independent variables. A logarithmic transformation was applied to the error scores in an attempt to normalise these positively skewed data (Tabachnick & Fidell, 2001). A significant overall effect was found only for sex ($F = 11.415, df = 1, p .001, \eta^2 = .029$), with males scoring significantly more errors than females. Table 4

TABLE 3

Standard Errors of Estimation (SE_e) and Prediction (SE_p) for the ELF Standard Scores

	SE_e				SE_p			
	Confidence Interval (%)				Confidence Interval (%)			
	68	90	95	99	68	90	95	99
ELF	5	8	10	13	8	13	16	21

Note: ELF = Excluded Letter Fluency Test.

TABLE 4

Base Rate of ELF Errors

Approx % of sample producing given or greater numbers of errors	Males	Females	Total
25	4	3	3
15	5	4	4
10	7	5	6
5	9	8	8
2	12	10	12
1	18	14	15

shows the percentage of males and females making particular numbers of errors on the ELF.

The correlation of total ELF raw scores with ELF error scores was not significant ($r = .064$, $p = .204$, $N = 399$).

Discussion

The aim of the present article was to present normative data for the ELF. As with initial-letter fluency tasks (e.g., Ruff et al., 1996) level of education was found to be influential in determining the number of words generated. There was no significant effect for gender or age on the total score. However, the age range of the sample was restricted to 18- to 34-year-old participants, which may account for the lack of effect of age. A significant effect for gender was found on the number of error scores that consisted of rule-breaks and perseverations, with males making more errors than females. This is an interesting finding, given this was a normal sample. Further research using the ELF as a self-monitoring measure in clinical samples will need to take this gender difference into account.

Previous studies have suggested a strong correlation between initial-letter fluency and estimates of intelligence; however, we found only a weak to moderate correlation between the ELF and a full version of the WAIS-R. Although this precluded the development of regression equations to predict ELF scores on the basis of intelligence level, it did establish that the ELF has discriminant validity. Future studies will be required to establish convergent validity as well as predictive validity.

This study also found weak to moderate correlations between the ELF and measures of reading ability. This suggests that low reading ability is not strongly correlated with ELF performance.

As Crawford et al. (1995) had shown that different fluency tasks were not differentially sensitive

to TBI, but that the ELF provoked significantly more error scores in the TBI sample, we have provided a table describing the base rate of errors in this study for both males and females. The finding of a lack of significant correlation between the total ELF score and the error score suggests that error scores are measuring a different process. This should assist future researchers in determining the value of ELF error scores as a measure of self-monitoring and evaluate its predictive validity in head trauma and other clinical samples.

References

- Atkinson, L., (1991). Three standard errors of measurement and the Wechsler Memory Scale-Revised. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3, 136–138.
- Beardsall, A., & Huppert, F.A. (1994). Improvement in NART word reading in demented and normal older persons using the Cambridge Contextual Reading Test. *Journal of Clinical and Experimental Neuropsychology*, 16, 232–242.
- Bechtoldt, H.P., Benton, A.L., & Fogel, M.L. (1962). An application of factor analysis in neuropsychology. *Psychological Record*, 12, 147–156.
- Benton, A.L. (1967). Problems of test construction in the field of aphasia. *Cortex*, 3, 32–58.
- Carstairs, J.R., & Shores, E.A. (2000). The Macquarie University Neuropsychological Normative Study (MUNNS): Rationale and methodology. *Australian Psychologist*, 35, 36–40.
- Crawford, J.R., Moore, J.W., & Cameron, I.M. (1992). Verbal fluency: A NART-based equation for the estimation of premorbid performance. *British Journal of Clinical Psychology*, 31, 327–329.
- Crawford, J.R., Wright, R., & Bate, A. (1995). Verbal, figural and ideational fluency in CHI (Abstract). *Journal of International Neuropsychological Society*, 1, 321.
- de Lemos, M.M. (1981). *WAIS-R Australian supplement*. Hawthorn: Australian Council for Educational Research.
- Dudek, F.J. (1979). The continuing misinterpretation of the standard error of measurement. *Psychological Bulletin*, 86, 335–337.
- Elfgren, C.I., Ryding, E., & Passant, U. (1996). Performance on neuropsychological tests related to single photon emission computerised tomography findings in frontotemporal dementia. *The British Journal of Psychiatry*, 169, 516–422.
- Hennessy, M., & Mackenzie, B. (1995). AUSNART: The development of an Australian version of the NART. In J. Fourez & N. Page (Eds.), *Treatment issues and long term outcomes: Proceedings of the 18th Annual Brain Impairment Conference, Hobart, Australia, 1994*. (pp. 183–188). Bowen Hills, QLD: Australian Academic Press.
- Lucas, S.K., Carstairs, J.R., & Shores, E.A. (2003). A comparison of methods to estimate premorbid intelligence

- in an Australian sample: Data from the Macquarie University Neuropsychological Normative Study (MUNNS). *Australian Psychologist*, 38, 227–237.
- Miller, E. (1994). Verbal fluency as a function of verbal intelligence and in relation to different types of pathology. *British Journal of Clinical Psychology*, 23, 53–57.
- Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.
- Ponsford, J.L., & Kinsella, G. (1992). Attentional deficits following closed head injury. *Journal of Clinical and Experimental Neuropsychology*, 14, 822–838.
- Ruff, R.M., Light, R.H., Parker, S.B. & Levin, H.S. (1996). Benton Controlled Oral Word Association Test: Reliability and updated norms. *Archives of Clinical Neuropsychology*, 11, 329–338.
- Spree, O., & Benton, A. L. (1969). *Neurosensory Centre Comprehensive Examination for Aphasia: Manual of directions*. Victoria, BC: Neuropsychology Laboratory, University of Victoria.
- Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics*. 4th Ed. Boston: Allyn and Bacon.
- Thurstone, L.L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised: Manual*. San Antonio: Psychological Corporation.

Appendix A

Administration Instructions for the Excluded Letter Fluency Test (ELF).

In this task, I am going to ask you to give me some words that do not contain a particular letter. In doing so, there are certain rules that I would like you to follow.

If I asked you to give me words without the letter ‘S’, you could not say ‘suitcase’, ‘mast’ or ‘poster’ because they have an ‘S’ in them.

Words like ‘table’, ‘drive’ or ‘clean’ are okay because they do not contain the letter ‘S’.

Another rule is that the words must have more than three letters. So, if the excluded letter was ‘S’, you could not say ‘pet’ or ‘fix’.

Also, you may not use proper nouns, that is, the names of people, places or brands. If the excluded letter was ‘S’, you could not say ‘Fiji’, ‘Peter’ or ‘Colgate’.

There is one more rule. You may not use the same word but with a different ending. So, if you had given me ‘drive’ you could not then say ‘driver’ or ‘driving’ and so on.

As a practice, please give me a few words that do not contain the letter ‘S’.

Okay, when I say the letter I want you give me as many words as you can think of that do not contain that letter and continue until I tell you to stop. Give me as many words as you can think of that do not contain the letter ‘A’.

Write down all responses for 90 seconds, then say:

Stop. Good, I will now give you another letter. Remember, words must be longer than 3 letters and not be names of brands, people or places.

Now, give me as many words as you can think of that do not contain the letter ‘E’.

Write down all responses for 90 seconds, then say:

Stop. Good, I will now give you another letter. Remember, words must be longer than 3 letters and not be the names of brands, people or places. Now, give me as many words as you can think of that do not contain the letter ‘I’.

Write down all responses for 90 seconds, then say:

Stop. Okay, that’s the end of this task.

After completing all three trials, ask the participants to spell any words which were rule-break errors to determine whether it was a monitoring error or simply a genuine spelling error. For example, they may think ‘beast’ does not have an ‘a’ on the ‘not a’ trial and erroneously consider that the word is spelt ‘beest’. If their spelling does not contain the designated letter it should not be treated as a rule-break error.