**Supplemental material**

***Factor analysis***

A factor analysis employing all measures of syntactic complexity, lexical diversity and fluency revealed a five-factor solution, accounting for 82.1% of the variance. Table S1 shows factor 1 with high loadings on both measures of lexical diversity, and factor 2 with high loadings on both measures of syntactic complexity. Factor 3 has high loadings on repetitions and mid-clause pauses, plus medium loadings on end-clause pauses and total silence. Factor 4 has high loadings on reformulations and false starts, while factor 5 has only one high loading, on replacements. Thus, factor 1 has to do with lexis, factor 2 with syntax, factors 3 and 4 with breakdown and repair fluency respectively, and factor 5, with one loading on a measure of repair fluency, is harder to label.

|  |
| --- |
| **Table S1** *Factor analysis for* Footballa |
|  |  |  | Component |  |  |
| Measure | 1 | 2 | 3 | 4 | 5 |
| Complexity  |  | .91 |  |  |  |
| Mean length of unit |  | .93 |  |  |  |
| VocD  | .94 |  |  |  |  |
| MTLD  | .93 |  |  |  |  |
| Reformulations  |  |  |  | .85 |  |
| Replacements  |  |  |  |  | .91 |
| False starts  |  |  |  | .89 |  |
| Repetitions  |  |  | .76 |  |  |
| Pauses mid clauses  |  |  | .85 |  |  |
| Pauses end clause  | -.37 |  | -.51 |  |  |
| Silence  |  | -.45 | .50 | -.40 | -.39 |
| Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.a |
| a. Rotation converged in 6 iterations. |

For *Picnic,* a correlation matrix revealed a variety of coefficients between .42 and .82. Kaiser-Meyer-Olkin’s value of sampling adequacy was .45, and Bartlett’s Test of Sphericity was significant at *p* < .001, confirming substantial correlation in the data. The factor analysis indicated a five-factor solution, accounting for 78.8% of the variance. Table S2 shows factor 1 with high loadings for both measures of lexical diversity, and a medium loading for lexical replacements. Factor 2 has high loadings on both measures of syntactic complexity. Factor 3 has high loadings on two measures of repair fluency: reformulations and false starts. Factor 4 has high loadings on repetitions and end-clause pauses, while factor 5 loads on two measures of breakdown fluency: mid-pause clauses and total of silence. Thus, factor 1 has to do with lexical choices, factor 2 with syntax, factor 3 with repair fluency, and factor 5 with breakdown fluency. Factor 4 clearly has to do with fluency, though it is harder to label more precisely than that.

|  |
| --- |
| **Table S2***Factor analysis for* Picnica |
| Measure | Component |
| 1 | 2 | 3 | 4 | 5 |
| Complexity  |  | .85 |  |  |  |
| Mean length of unit  |  | .94 |  |  |  |
| VocD P | .90 |  |  |  |  |
| MTLD | .85 |  |  |  |  |
| Reformulations  |  |  | .81 | -.31 |  |
| Replacements  | .59 |  |  |  |  |
| False starts  |  |  | .91 |  |  |
| Repetitions  |  |  |  | -.78 |  |
| Pauses mid clause  |  |  |  |  | .85 |
| Pauses end clause  |  |  |  | .85 |  |
| Silence total |  |  | -.33 | .32 | .77 |
| Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalizationa |
| a. Rotation converged in 6 iterations. |

In sum, the factor analyses indicate that syntactic complexity, lexical diversity and fluency are distinct constructs, with fluency appearing to have a more complex structure than a simple division into its repair and breakdown aspects.

***Verification of statistical assumptions for t-test analysis***

**Table S3**

*Assumptions of the t-test*

|  |  |  |
| --- | --- | --- |
| *t*-test | Assumption | How corroborated |
| Sampling | The participants in the study were randomly sampled from the source population | Invitations to participate in the study were extended to all students at the Te Pūkenga site who met the inclusion criteria. They were drawn from seven different classes. |
| Sample size | The sample size calculated for the study is achieved | Statistical power was calculated using G\*Power (Faul et al., 2007). For matched pairs with an α error probability set at 0.05 and a medium effect size *dz* (0.5), an actual power of 0.95 was generated by a participant sample size of 45. Each candidate who volunteered to participate was recruited, resulting in a total number of 51 participants who completed the tasks. |
| Normal distribution | The scale of measurement of outcome variables are continuous and have an approximate normal distribution.  | One outlier was identified for one of the fluency measures (total silence) and was removed.For the total silence measure, for 50 participants the difference between their Football and Picnic performances ranged between -8.1s and +8.9s. Of any participant, the outlier had both the lowest amount of silence for Football and the highest for Picnic, resulting in a total difference of -38.8s of silence. This was deemed to represent unexplained ‘noise’ and was removed from the data.Visual inspection of the histograms and Q-Q plots indicated approximate normal distribution for each of the outcome variables. For each output variable, the results of Kolmogorov-Smirnov tests indicated that distributions were not significantly different. |