

# Supplementary Material

## Assessment of Language Function in Older Mandarin-Speaking Adults with Mild Cognitive Impairment using Multifaceted Language Tests

### Construction, Administration, and Scoring of the Mandarin Chinese Language Test

#### *Oral production domain*

The phoneme, word, sentence, and discourse levels of oral expression were evaluated. The oral production domain was composed of four tests: (1) the Zhuyin blending and deletion test, (2) sentence production and spontaneous speech test, (3) noun and verb naming test, and (4) verbal fluency test. Each test is described below in details.

#### Zhuyin blending and deletion test

This test examined participants' ability to use and pronounce Zhuyin symbols. The test consisted of two sections: Zhuyin blending and Zhuyin deletion. Each section contained eight items, including four real syllables (e.g., *ben1* ㄅㄣˊ) and four pseudo syllables (e.g., *piang1* ㄆㄧㄤˊ ㄟ), which were further classified on the basis of the number of Zhuyin symbols, with two two-Zhuyin symbols (e.g., *shan1* ㄕㄢˊ) and two three-Zhuyin symbols (e.g., *xian1* ㄒㄩㄢˊ). In the Zhuyin blending section, participants were asked to read monosyllabic Zhuyin symbols presented in written form. In the Zhuyin deletion section, participants were asked to delete the first Zhuyin symbols (e.g., *shi1* ㄕㄟˊ) from auditory syllables (e.g., *shan1* ㄕㄢˊ) and pronounce the remaining phonetic blends (e.g., *an1* ㄢˊ).

Two points were assigned for each correct response provided within 10 s for each item. One point was assigned if a response was self-corrected, delayed by more than 10 s, or was provided

after a replay of the auditory stimulus. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer. The total score ranged from 0 to 16 points in each section.

#### Sentence production and spontaneous speech test

This test evaluated participants' ability to verbally produce sentences or connected speech by describing pictures. The test consisted of two sections: sentence production and spontaneous speech. The sentence production section included five pictures depicting constrained subjects, objects, and verbs. A constrained subject was indicated by a star symbol near an arrow for each picture, whereas a constrained object was indicated by an arrow. A written verb was provided in each picture. For example, in the picture of the item "A groom is hugging a bride," the groom was indicated by a star near an arrow and the bride was indicated by an arrow. The word "hug" was provided under the drawing. Each picture had different syntactic structure requirements, including active, passive, and dative sentences. Participants were asked to orally produce a complete sentence for each picture. Each sentence was scored on the basis of three components: vocabulary, thematic role, and syntactic structure. The highest possible score in each component ranged from 1 to 3 points on the basis of the predetermined correctness of answers. The highest possible score of each item ranged from 5 to 7 points on the basis of the sum of scores of the three scoring components. The total score of the section ranged from 0 to 28 points.

The spontaneous speech section included six questions on personal information (e.g., occupation) and experiences (e.g., How was your day yesterday?) and a picture depicting a park scene with information units of 4 people, 6 actions, and 16 objects. Participants were first asked to answer the six questions and were then instructed to describe the content of the picture within 3 min. Scoring of spontaneous speech was based on two linguistic elements: information content

and fluency. Each element was scored from 0 to 10 points on the basis of the number of information units and speech quality (see Supplementary Table 1). The total score of both the elements combined ranged from 0 to 20 points.

### Noun and verb naming test

This test evaluated participants' ability to produce the names of objects and actions on pictures. The test included 16 nouns in four categories (four animals, four vegetables and fruits, five clothing, and three tools) and 16 verbs in two categories (10 transitive verbs and 6 intransitive verbs). The familiarity rating scores did not significantly differ across the four categories for the noun items (animals:  $M = 4.67$ ,  $SD = 0.77$ ; vegetables and fruits:  $M = 4.94$ ,  $SD = 0.25$ ; clothing:  $M = 4.88$ ,  $SD = 0.37$ ; tools:  $M = 4.94$ ,  $SD = 0.18$ ) or verb items (transitive:  $M = 4.71$ ,  $SD = 0.65$ ; intransitive:  $M = 4.67$ ,  $SD = 0.66$ ). Participants were asked to name pictures depicting an object or an action. The noun naming section always preceded the verb naming section.

Two points were assigned for each correct response provided within 10 s. One point was assigned if a response was partially correct, self-corrected, or delayed by more than 10 s. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer. The total score ranged from 0 to 32 points in each section. Moreover, naming errors for noun and verb items were categorized into semantic or phonological errors on the basis of criteria indicated by Druks et al. [1]. A semantic error was defined as a semantically related response to the target word (e.g., naming dogs as cats). A phonological error was defined as a response that included similar phonemes but was semantically unrelated to the target word (e.g., naming 毛巾 *mao2 jin1* as 毛肩 *mao2 jian1*).

Furthermore, a semantic error was classified into four subtypes: coordinate, superordinate, subordinate, and associate. A coordinate error constituted a response within the same category as the target word (e.g., “dog” for cat). A superordinate error involved a response belonging to a more general category than that of the target word (e.g., “pet” for cat). A subordinate error encompassed a response belonging to a narrower category than that the target word (e.g., “British shorthair” for cat). An associate error entailed a response not belonging to the same category as the target word but showing certain relations (e.g., “scratchers” for cat).

### Verbal fluency test

This test examined participants’ ability to generate as many exemplars as possible on the basis of a rule within a time constraint. The test was composed of two sections: single category and category switching. The single category section contained three semantic categories (i.e., animals, fruits, and appliances) that varied in animacy. The category switching section included two categories (i.e., vegetables and vehicles). In the single category section, participants were asked to provide as many exemplars as possible within 1 min. In the category switching section, participants were instructed to provide as many exemplars as possible within 1 min but were asked to alternate between vegetables and vehicles.

Regarding scoring in both the sections, 1 point was assigned for each correct response, including dialectal variations. Intrusive responses from another category were not assigned any point. Repetitions of an exemplar were counted as only one correct response. Any names of subcategories were not scored (e.g., “dog” under the animal category) if specific items within that category (e.g., poodle and golden retriever) were also mentioned. The total test score was the sum of points received for the four test items (i.e., three single categories and one category switching).

### *Auditory and reading comprehension domain*

The phoneme, tone, word, and sentence levels of the comprehension ability were assessed using seven tests: (1) the phoneme and tone recognition test, (2) word–picture matching test, (3) attribute verification test, (4) relatedness judgement test for concrete and abstract words, (5) sentence and syntactic comprehension test, (6) idiom comprehension test, and (7) written lexical decision test. Each is described below.

#### Phoneme and tone recognition test

This test evaluated participants' ability to discriminate monosyllabic auditory pairs in same and different tasks and to identify lexical tones in a matching task. This test involved two tasks: the phoneme and tone discrimination task and the tone-matching task. In the phoneme and tone discrimination task, 19 pairs of monosyllabic auditory words were used, with the same (eight pairs) or different (11 pairs) Zhuyin symbols or tones in each item. Items varied in the number of Zhuyin symbols between stimuli (one symbol: six items; two symbols: seven items; three symbols: six items). Nine pairs out of the 19 items were used for discriminating Zhuyin symbols (tone 1: one item; tone 2: three items; tone 3: two items; tone 4: three items), and 10 pairs were used for discriminating tones (with all possible paired combinations of four tones). Participants were asked to orally respond either “the same” or “different” for the 19 pairs of Zhuyin symbols or tones.

In the tone-matching task, eight monosyllabic words were included, with four real syllables and four pseudo syllables that consisted of four tones in both types of syllables. After listening to each auditory stimulus, participants were asked to choose the best answer from four written tonal signs.

For the two tasks, 2 points were assigned for each correct response provided within 10 s. One point was assigned if a response was self-corrected, delayed by more than 10 s, or was provided after a replay of the auditory stimulus. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer. The total score ranged from 0 to 38 points in the phoneme and tone discrimination task and from 0 to 16 points in the tone matching task.

### Word–picture matching test

This test examined participants' ability to match the meanings of a word, either in auditory or written form, to its corresponding picture. Fourteen concrete nouns varying between one to two characters had a high level of word frequency (the first one-third of frequency ranking) [2]. The 14 nouns were distributed among three semantic categories (four animals, five vegetables and fruits, and five tools). The familiarity scores did not significantly differ for the items belonging to the three categories (animals:  $M = 4.80$ ,  $SD = 0.45$ ; vegetables and fruits:  $M = 4.96$ ,  $SD = 0.11$ ; tools:  $M = 4.95$ ,  $SD = 0.14$ ). Each target was paired with three types of foil pictures that were semantically related, phonologically related, and unrelated to targets. Participants were asked to choose one picture that best matched the meanings of target nouns. The examiner presented 14 items auditorily or visually above the four picture stimuli. The written form was always administered before the auditory form. Two points were assigned for each correct response provided within 10 s. One point was assigned if a response was self-corrected, delayed by more than 10 s, or was provided after a replay of the auditory stimulus. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer. The total score ranged from 0 to 28 points in each auditory and written form.

### Attribute verification test

This test evaluated participants' ability to process the semantic attributes of concrete nouns through a sentence verification task in auditory and written forms. Ten concrete nouns contained a moderate to high level of word frequency (the first two-thirds of frequency ranking) [2] and were distributed among four semantic categories (two animals, three vegetables, two fruits, and three tools). Items of the four categories did not differ significantly in the familiarity rating score (animals:  $M = 4.59$ ,  $SD = 0.80$ ; vegetables:  $M = 4.96$ ,  $SD = 0.17$ ; fruits:  $M = 4.84$ ,  $SD = 0.44$ ; tools:  $M = 4.85$ ,  $SD = 0.32$ ) but differed in the concreteness rating score,  $F(1.17, 17.59) = 5.51$ ,  $p = 0.026$ ,  $\eta_p^2 = 0.269$ . Planned paired  $t$  tests indicated that items belonging to the animal category had significantly lower concreteness scores than did those belonging to the vegetable category,  $t(15) = -2.45$ ,  $p = 0.027$ ,  $d = -0.612$ , fruits,  $t(15) = -2.35$ ,  $p = 0.033$ ,  $d = -0.588$ , and tools,  $t(15) = -2.45$ ,  $p = 0.027$ ,  $d = -0.612$ .

For each concrete noun (e.g., watermelon), participants were asked to provide a “yes” or “no” response to two statements related to the noun (e.g., “Watermelon can be eaten”) and to choose the better answer from two choices of a question (e.g., “Is the shape of watermelon round or diamond?”). All the stimuli were displayed auditorily or in written sentences in a fixed order. The 30 sentences, which were distributed to 10 noun stimuli, varied in three semantic relations (i.e., functional, perceptual, and part/whole) and veracity of statements (10 factual and 10 false). The written form was always administered before the auditory form. Two points were assigned for each correct response provided within 10 s. One point was assigned if a response was self-corrected, delayed by more than 10 s, or was provided after a replay of the auditory stimulus. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer. The total score ranged from 0 to 60 points in each auditory and written form.

### Relatedness judgement test for concrete and abstract words

This test examined participants' ability to identify semantic relations between concrete or abstract two-character words in auditory and written forms. The test included 20 items for each form, and each item had one target and two choices of words. Participants were instructed to indicate which choice was semantically related to the target word. A target word was read before two choice words by the examiner in the auditory form or was displayed above two choice words on paper in the written form.

Target words were distributed among nouns (eight items), verbs (eight items), adjectives (two items), and adverbs (two items). Both the noun and verb subtypes had four high-frequency words and four low-frequency words, which were further classified on the basis of imageability with two words with low-to-mediate imageability and two words with high imageability [3]. The adjective and adverb items were self-selected with a high level of frequency (the first one-third of frequency ranking) [2]. All the items had high levels of familiarity (4.06 to 4.88 on a 5-point rating scale) and concreteness rating (3.75 to 4.94 on a 5-point rating scale) scores. The written form was always administered before the auditory form.

Two points were assigned for each correct response provided within 10 s. One point was assigned if a response was self-corrected, delayed by more than 10 s, or was provided after a replay of the auditory stimulus. Zero points were assigned when participants failed to respond or provided an incorrect or unrelated response. The total score ranged from 0 to 40 points in each form.

### Sentence and syntactic comprehension test

This examined participants' ability to match the meaning of a sentence, either in the auditory or written form, to its corresponding image. Seventeen sentences with 5 to 12 characters were

varied in sentence types, such as active, passive, and 把 *ba 3* (e.g., “女孩 <1> 「把」 父親 <2> 叫起床 <3> The daughter <1> wake up <3> her father <2>”) sentence structures [4]. Stimuli contained six semantically irreversible sentences and 11 reversible sentences. For each question, one target picture was paired with three foil pictures representing false sentences with a substituted subject, verb, or object of the target sentence. For example, for the target sentence “A cook is chasing a mail carrier,” the three foils were “A mail carrier is chasing a cook,” “A cook is hitting a mail carrier,” and “A mail carrier is hitting a cook.”

Participants were asked to choose one picture that best matched the meanings of the sentences. All the 17 target sentences were presented auditorily by the examiner or visually shown above the four picture stimuli. Both the auditory and written forms were parallel versions using the same pictures but different target items. Each target sentence in one form was substituted by a foil sentence in another form. Two points were assigned for each correct response provided within 10 s. One point was assigned if a response was self-corrected, delayed by more than 10 s, or was provided after a replay of the auditory stimulus. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer. The total score ranged from 0 to 34 points in each auditory and written form.

### Idiom comprehension test

This test examined participants’ ability to explain the nonliteral meanings of idioms stored in semantic knowledge. Eight idioms varying in four to seven characters had high familiarity ratings (4.63 to 4.88 on a 5-point rating scale) and concreteness ratings (4.00 to 4.44 on a 5-point rating scale). Participants were asked to verbally explain each auditorily presented idiom for its nonliteral or deep level of meaning. Two points were assigned for each correct response provided

within 10 s. One point was assigned if an idiom was explained with only literal meaning or not entirely correct. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer. The total test score on the eight items ranged from 0 to 16 points.

### Written lexical decision test

This test examined participants' ability to identify the valid lexical forms of two-character words presented visually. The test contained 16 items from the study conducted by Lin [5], with eight real words and eight pseudo words. All the real words had high levels of word frequency (the first one-third of frequency ranking) [2] and familiarity rating scores (3.94 to 5.00 on a 5-point rating scale). Participants were asked to respond either "yes" or "no" to each item to indicate whether it is a valid word. Two points were assigned for each correct response provided within 10 s. One point was assigned if a response was self-corrected or delayed by more than 10 s. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer. The total test score on the 16 items ranged from 0 to 32 points.

### *Reading aloud domain*

Three tests used to assess the word to sentence levels of the reading aloud ability were included: the word reading test (for regular and irregular words), the nonword reading test, and the sentence reading test. Each test is described below in details.

### Word reading test

This test examined participants' ability to recognize and pronounce regular and irregular characters presented visually. A regular word is defined as the articulation of the character (e.g.,

試 *shi4*) identical to its phonetic radical (e.g., 式 *shi4*). By contrast, an irregular word is defined as a different articulation between the character (e.g., 抽 *chou1*) and its phonetic radical (e.g., 由 *you2*). The test contained 10 regular and 10 irregular characters, including four characters without phonetic radicals (e.g., 傘 *san3*). The familiarity rating scores did not differ significantly between regular and irregular characters (regular:  $M = 4.42$ ,  $SD = 0.95$ ; irregular:  $M = 4.58$ ,  $SD = 0.82$ ). Each type of stimuli contained six high-frequency words (the first one-third of ranking) [2] and four low-to-moderate-frequency words (the last two-thirds of ranking). The 10 regular characters were classified on the basis of consistency, with three consistent and three inconsistent characters in the high-frequency group, and two consistent and two inconsistent characters in the low-frequency group. Consistent characters referred to characters sharing the same phonetic radicals and pronunciation, whereas inconsistent characters are those pronounced differently despite sharing the same phonetic radicals. Participants were asked to read aloud each character visually presented to them.

The test contained 20 items in total. Two points were assigned for each correct response provided within 10 s. One point was assigned if a response was self-corrected or delayed by more than 10 s. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer. Reading errors were classified into three types: semantic, regularizing, or unspecified [6]. A semantic error was a substitute response semantically related to the target word. A regularizing error was a character that was falsely pronounced by relying on the phonetic radical, such as reading 迪 *di2* as 由 *you2*. An unspecified error referred to an incorrect response that did not belong to a semantic or a

regularizing error, such as an orthographic-related error (e.g., reading 娛 *yu2* as 誤 *wu4*). The total test score ranged from 0 to 20 points for each section of regular and irregular characters.

#### Nonword reading test

This test examined participants' ability to read aloud noncharacters through character's phonological information (i.e., phonetic radicals) [7]. Four noncharacters were included in this test. Two consistent and two inconsistent stimuli were classified on the basis of the number of orthographical neighbors (ONs). An ON was defined as a set of characters with the same phonetic radical neighbors. In both the consistent and inconsistent types of noncharacters, one noncharacter had a high number of ONs (>10), whereas another had a low number of ONs (<10). Participants were asked to read aloud each noncharacter presented visually. The correct answers for each noncharacter were predetermined as all pronunciations from its ONs.

Two points were assigned for each correct response provided within 10 s. One point was assigned if a response was self-corrected or delayed by more than 10 s. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer. The total test score ranged from 0 to 8 points.

#### Sentence reading test

This test examined participants' ability to recognize and pronounce sentences that were presented visually. Ten sentences, ranging from 3 to 12 characters, were used in the test. These sentences varied in sentence types, such as active, passive, and question sentence structures [4]. Participants were asked to read aloud each sentence. The test was always administered after the sentence repetition test due to their shared stimuli.

Two points were assigned for each correct response provided within 10 s. One point was assigned if a response was self-corrected or delayed by more than 10 s. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer. Reading errors were classified into one of the four types: omission, semantic, phonological, and unspecified. An omission error indicated a few characters were omitted from the target sentence. A semantic error indicated a response using a semantically related substitute to the target sentence. A phonological error was a substitute response that was phonologically similar to the target word. An unspecified error indicated the incorrect response that did not belong to an omission, semantic, or phonological error type. The total test score ranged from 0 to 20 points.

### *Repetition domain*

Two tests were included, namely the word and nonword repetition test and sentence repetition test, to determine the phoneme, word, and sentence levels of the repetition ability. Details of both the tests are provided below.

#### Word and nonword repetition test

This test examined participants' ability to repeat words and nonwords. The section on word repetition contained 10 items, with two monosyllabic Zhuyin symbols (a one-Zhuyin item and a two-Zhuyin item) and 8 two-character words. The 8 two-character words had high word frequencies (the first one-third of ranking) [2], familiarity ratings (4.38 to 4.88 on a 5-point rating scale), and concreteness ratings (3.75 to 5.00 on a 5-point rating scale). Participants were asked to repeat 10 items that were presented auditorily. In the section on nonword repetition, five items consisting of two to six syllables that were formed from low-frequency characters were

used according to the study conducted by Chen and Liu [8]. Participants were asked to repeat five nonwords that were auditorily presented.

In both the word repetition and nonword repetition sections, two points were assigned for each correct response provided within 10 s. One point was assigned if a response was self-corrected, delayed by more than 10 s, or was provided after a replay of the auditory stimulus. Zero points were assigned when participants failed to provide a response or provided an incorrect or unrelated answer.

Errors were classified into one of the four types: dysarthria, phonological, semantic, and unspecified. Dysarthria is a motor speech disorder causing muscle weakness and difficulties in articulating clearly. A phonological error was a response with additional or omitted phonemes or with different tone features. A semantic error was a substitute response that was semantically related to the target sentence. An unspecified error was a response that could not be classified as any of the other three types. The total score ranged from 0 to 20 points in the word repetition and from 0 to 10 points in the nonword repetition.

### Sentence repetition test

This test examined participants' ability to repeat sentences. The test contained 10 sentences in total. The test materials, scoring procedures, and error classification were the same as those used for the sentence reading test. The examiner read sentences for participants to repeat. The test was always administered before the sentence reading test because of the use of the same materials. The total score ranged from 0 to 20 points.

### *Writing domain*

The word to sentence levels of the writing ability were examined using four tests: (1) the written naming test, (2) written to dictation test, (3) written sentence test, and (4) copy of the written word test. Details of each test are provided below.

#### Written naming test

This test examined participants' ability to write the names of objects presented in pictures. Five nouns ranging from one to three characters were included, with a total of 10 characters in all five nouns. The five words contained different regular and irregular characters. All words had high levels of word frequency (the first one-third of frequency ranking) [2] and familiarity rating scores (4.88 to 5.00 on a 5-point rating scale). Participants were asked to write the names of objects depicted in pictures.

Items were scored on the basis of the correctness of each character. Simplified or scribbled characters were not allowed. Each character was assigned 1 point if it was written correctly. If any error was made, the score of each character was obtained by subtracting 0.5 from 1 accumulatively until a score of 0 was achieved. Writing errors were classified into four types: unintelligible words, orthographic errors, semantic errors, and unspecified errors. An unintelligible word referred to a character containing additional or omitted strokes (i.e., basic signs in a Chinese character). An orthographic error referred to a character that was a homophone of the target character or that had a substitute or misplaced radical. A semantic error referred to a replacement that was semantically related to the target sentence. An unspecified error referred to a response that did not belong to the mentioned error types, such as writing with only Zhuyin symbols. The total score ranged from 0 to 10 points for the 10 characters.

### Written to dictation test

This test examined participants' ability to write words that were presented auditorily. Five nouns ranging from two to three characters were included, with a total of 11 characters. The five words contained different numbers of regular and irregular characters. All words had high levels of word frequency (the first one-third of frequency ranking) [2] as well as had high familiarity rating scores (4.25 to 5.00 on a 5-point rating scale) and concreteness rating scores (3.81 to 5.00 on a 5-point rating scale). Participants were asked to write the words the examiner read. Scoring and error classification were the same as those used in the written naming test. A total of 0.5 points were subtracted if participants asked the examiner to reread an item without apparently having a hearing or concentration problem. The total score ranged from 0 to 11 points for 11 characters.

### Written sentence test

This test assessed participants' ability to write basic sentence structures on the basis of information provided in pictures. Five sentences were included in the test. The test materials and scoring method were the same as those used in the sentence production section. Participants were asked to write a complete sentence for a constrained subject, a verb, and objects shown on a picture. Participants' writing with some additional or omitted strokes of the recognizable Chinese character was considered acceptable. The total score ranged from 0 to 28 points for five sentences.

### Copy of the written word test

This test examined participants' ability to percept and construct written words and nonwords. Five items ranging from one to two characters were included in the test, with four real

words and one noncharacter. The five items varied in stroke directions and numbers. The four real word items had a high level of familiarity rating scores (4.31 to 4.63 on a 5-point rating scale). Participants were asked to copy each word precisely.

Five items were scored on the basis of the correctness of each stroke. Each stroke received 1 point if it was written correctly. Copy errors were classified into two types: direction and stroke. A direction error referred to a stroke with a deviated orientation that was difficult to recognize as the original stroke. A stroke error referred to the addition or omission of a stroke from a character. A point of 0 was assigned if a direction error was made. In the case of a stroke error, 1 point was subtracted for each character accumulatively until a score of 0 was reached. The highest possible score of each item was the number of strokes in an item. The total score ranged from 0 to 101 points.

## REFERENCES

- [1] Druks J, Masterson J, Kopelman M, Clare L, Rose A, Rai G (2006) Is action naming better preserved (than object naming) in Alzheimer's disease and why should we ask? *Brain Lang* **98**, 332-340.
- [2] Cheng C-C, Huang C-R, Lo F-J, Tsai M-C, Hwang Y-C, Chen X-Y, Lu C-J, Han Y-C, Li C-C, Digital Resources Center for Global Chinese Teaching and Learning, Institute of Linguistics, Academia Sinica, [https://elearning.ling.sinica.edu.tw/eng\\_teaching.html](https://elearning.ling.sinica.edu.tw/eng_teaching.html), Accessed February 19, 2019.
- [3] Hu J-F, Chen Y-C, Zhuo S-L, Chen H-C, Chang Y-L, Sung Y-T (2017) Word Association Norms and Associated Responses: Reference index for 1200 two-character Chinese words. *Bull Educ Psychol* **49**, 137-161.

- [4] Lin B-G, Huang R-J (1999) Indices of written language assessment in Chinese children: Syntax and semantics. *Bull Special Educ* **17**, 163-188.
- [5] Lin W-C (2007) The Role of Phonological Information in Chinese Compound Word Recognition. Thesis. National Taiwan University.
- [6] Staffaroni AM, Weintraub S, Rascovsky K, Rankin KP, Taylor J, Fields JA, Casaletto KB, Hillis AE, Lukic S, Gorno-Tempini ML (2021) Uniform data set language measures for bvFTD and PPA diagnosis and monitoring. *Alzheimers Dement (Amst)* **13**, e12148.
- [7] Lin Y-F (2011) Consistency and regularity effects of naming Chinese character in alexic patients. Thesis. Department of Speech Language Pathology and Audiology, National Taipei University of Nursing and Health Sciences.
- [8] Chen Y-C, Liu H-M (2018) Nonword repetition performance in Mandarin-speaking preschool children with specific language impairment—a study of diagnostic accuracy. *Bull Special Educ* **43**, 57-81.

**Supplementary Table 1.** Scoring table for the spontaneous speech section

Score	Speech characteristics
<i>Information content</i>	
0	No response or expression
1	Only incomplete responses (e.g., parts of the name or the address were given)
2	One correct answer for six simple questions
3	Two correct answers for six simple questions
4	Three correct answers for six simple questions
5	Three correct answers for six simple questions and a few descriptions regarding the picture
6	Four correct answers for six simple questions and a few descriptions regarding the picture
7	Four correct answers for six simple questions and the picture descriptions that contained at least six items of people, actions, and objects in sum
8	Five correct answers for six simple questions but incomplete descriptions regarding the picture (phonemic paraphasia was considered as a correct response if it was recognizable by the examiner)
9	Total correct answers for 6 simple questions and at least 10 statements to the pictures, including the people, actions, and objects
10	Total correct answers for six simple questions and complete picture descriptions (length and complexity of sentences were within normal limits, and almost all the items in the picture were mentioned)
<i>Fluency</i>	
0	Responses with only short and meaningless articulation
1	Repetitive articulation with tonal differences that seemed to express meanings
2	Responses with only single words, often accompanied with paraphasia and effortful, and halting speech
3	Fluent but repetitive utterances or mumble
4	Halting or telegraphic speech often with single words and paraphasia; spontaneously short and limited utterances when answering, such as “I don’t know”
5	Expression with a few syntactic structures and slight fluency but often in a telegraphic style
6	Expression with nearly complete and normal syntactic structures but with occasional paraphasia

- 
- 7 Fluent expression with neologisms and phonemic jargon that did not violate the phonological rules of Mandarin Chinese and were phonemically related to target words

---

  - 8 Fluent but circumlocutory expressions with obvious word-finding difficulty, verbal paraphasia, and possible semantic jargon (i.e., substitute words that were semantically related to target words); complete but unrelated sentences

---

  - 9 Nearly comprehensive and related responses with possible pauses, paraphasia, word-finding difficulty, and articulation errors occasionally

---

  - 10 Expression with normal level of sentence length and complexity, without an obvious speed problem, pauses, and articulation errors or paraphasia
-