The Differences in English Sentence Repetition Performance between Chinese High School Students at Risk for Language Disorder and Typically-developing Peers

Luning Zhang[1*]

[1]School of English and International Studies, Beijing Foreign Studies University, 10080, Beijing, China

Abstract. Developmental Language Disorder (DLD) causes harms to patients and the risk for getting DLD can be seen in many students. In Chinese exam- oriented education, students are facing enormous difficulties in L2 learning because of the risk of developmental language disorder, which warrants research. Previous studies mostly use Sentence Repetition Task (SRT) as a clinical marker and seldom seldom examine aspects such as error types and syntactic breakdowns, not to mention Chinese high school context. This research aims at finding the differences in English SRT performance between Chinese high school students at risk and typically-developing peers, revealing the predictive factors of SRT performance. It was done in Zhejiang province, applying questionnaire and SRT. The finding is that the at-risk group perform significantly poorer than TD group, and the higher the syntactic complexity of the task, the more obvious differences are. The risk for speaking disorder shows the strongest predictive power while the reading disorder risk shows the weakest.

1 Introduction

Developmental Language Disorder (DLD) is a neurodevelopmental condition which can cause various difficulties in acquiring and using languages. According to previous researches, DLD is reported to affect more than 7% of the population around the globe [1]. DLD can happen in childhood, puberty and adulthood, with obvious symptoms such as difficult struggles with complex syntax, extremely short verbal working memory, and the processing of lengthy auditory information [2]. In school setting, not only can DLD and the tendency of DLD lead to living difficulties, but it can also cause severe academic barriers in all disciplines [3].

With the development of globalization, second language acquisition (SLA) is gaining increasing popularity worldwide, which has already been included in many high school curriculums and has become one of the contents of college entrance examination. The process of SLA involves instructed courses, drills and memorizations, which calls for students'

* Corresponding author: ZhangLuning@hhu.edu.cn

phonological working memory, auditory discrimination, and procedural memory [4]. As mentioned above, those students with high DLD risk can be extremely vulnerable, for their inability in memorizing words in a short period of time, failure in mimicking phonetic sounds, and so forth. This critical gap between students at risk for DLD and typically-developing (TD) peers needs to be fully aware, and the detailed difficulties which students at risk for DLD face need to be clearly revealed for pedagogical purposes and for these vulnerable students' future development.

Therefore, this study aims to further contribute to the study of DLD by evaluating the detailed performance of Chinese high school students at risk for DLD on every step of an English L2 Sentence Repetition Task (SRT), which is widely acknowledged as a reliable method to diagnostically robust clinical marker for DLD across various languages and ages [5].

2 Literature review

2.1 The centrality of language processing

Generally speaking, DLD is considered as a negative influencing factor in the cognitive process of second language acquisition in most of the previous researches. The cognitive processing framework, which emphasizes memory and real-time decoding, is distinct from traditional approaches of SLA, which measure learners' L2 performance by emphasizing specific grammatical rules, rather, it focuses on the integration of a wide range of cognitive components, which includes phonological short-term memory, syntactic parsing, semantic integration, etc. [6]. In an attempt to measure these cognitive performances, the sentence repetition task (SRT) was proposed. Not only can SRT test learners' short-term memory, but it also asks learners to actively perceive and distinguish a certain acoustic signal, parsing its' syntactic structure, decoding the underlying semantic meaning and finally phonologically repeat it. The common symptoms of DLD include having difficulties in interpreting acoustic signals, failing in organizing sentences or pronouncing sentences and so forth, which can cause the failure in different stages of SRT tasks [7]. According to the processing framework, SRT can successfully tap into the core processing weaknesses that was brought by DLD.

2.2 SRT as a clinical marker in L1 and L2 contexts

In the past few years, SRT has been widely acknowledged as an accurate clinical marker of DLD around the globe, especially in L1 acquisition, for its comprehensive evaluation of learners' language ability. For example, Praat once conducted studies about DLD's influence on bilinguals' SRT performance with and without DLD. Their study revealed that general bilingual participants from all age groups with DLD perform significantly worse on SRTs than TD peers in aspects of memory, vocabulary, and exposure [8]. Also, Taha did research among 30 Arabic children with DLD and 60 TD children who are of the same age and finally found out that SRT coring methods showed good diagnostic power in distinguishing DLD kids from the normal kids [9].

In the realm of L2 acquisition, SRT is also used predominantly as a tool to differentiate between TD bilingual children and monolingual children with DLD. The work of Scheidnes and Tuller revealed that the performance in manipulating syntactic complexity in SRT can differentiate children with DLD and TD peers [10]. Also, Gao and Ma conducted a study about L2 grammatical learning and they applied SRT as part of their assessment battery to measure participants' language level, which showcased high reliability in evaluating learners' general L2 ability contemporarily [11].

2.3 Identified research niche

Previous researches have reached common grounds: First, DLD obviously brings negative effects to L2 learning. Second, the differences in L2 learning between learners with DLD and their TD peers can be clearly reflected in SRT.

However, despite these agreements, there exists some research gaps. To begin with, SRT is mostly used as a clinical tool to identify DLD in L1 and L2 contexts, differentiating people at risk of DLD from TD peers, but the detailed research concerning the differences in SRT performance in L2 contexts remains scarce [10]. And then, the Chinese high school context is of great importance. Since most previous researches are done in the western world, the research concerning the effects of DLD risks on SLA in China is still lacking, despite its exam-oriented education system and high-stakes language testing. In Chinese high school, where the exam-oriented education is applied, students are immersed in a highly competitive environment driven by national college entrance exam and intensive daily language instruction, which may add extra effects on their DLD symptoms and thus affect their response and L2 performance.

Therefore, this study is designed to address the gaps by investigating in detail the different performance of Chinese high school students at risk for DLD and their TD peers on an English L2 SRT. There are two research questions:

RQ1: Are there any differences in English Sentence Repetition Performance between Chinese high school students at risk for language disorder and typically-developing Peers? If the differences exist, what are they?

RQ2: What are the predictive factors in performance of SRT?

3 Methodology

3.1 Participants

The research is conducted in a normal high school in the south east part of China, where the economy is relatively well-developed and English education covered almost all the students. 200 participants from grade 12 are randomly selected from 700 12th grade students. 100 of them are females, and 100 of them are males.

Relatively speaking, the overall English grades of this high school are above average in China for the high economic development level of Zhejiang province. 5 English classes are provided to students per week, one class per day, for 40 minutes. 3 out of the 5 English classes are based on the text materials and English listening, reading, speaking, and writing knowledge are taught according to this standardized material. The other 2 classes focus on the feedback of students' homework. The homework is the comprehensive test papers covering vocabulary, grammar, reading comprehension, and writing tasks based on the teaching content.

3.2 Instruments

3.2.1 The questionnaire

In order to distinguish students at risk for language disorder and typically developing peers, the questionnaire is adapted from Adult Developmental Coordination Disorders Checklist (ADC). It is a validated questionnaire tool for diagnosing developmental disorder which is applied widely in previous studies, covering a wide range of main areas of daily activity performance in which adults with or in the risk of getting DCD characteristically experience

systematic barriers, ranging from language disorder, attention, to self-care [12]. In previous studies, ADC is often used as a 5-Likert scale (zero means strongly disagree and 5 stands for strongly agree) and a total score was calculated for each participant, former researcher got used to assigning top 25% of participants as at-risk group, while assigning the bottom 25% of participants as TD group, not taking the mid 50% percent into consideration [13]. Kirby has tested the validity of ADC. The Alpha coefficient obtained from the 107 questionnaires shows that the value for the ADC, which was determined to be 0.953 is well above the level of acceptability, determining great reliability [12].

In this study, the author picked the items that are related to language disorder in ADC as bases and adapted them to fit into the research's need, and translated into the Chinese version, comprehensively evaluated students' difficulty in language using, including 4 dimensions: listening speaking, reading and writing. Meanwhile, the author followed the former approach of dividing the groups according to the questionnaire results, with 50 of them defined as at-risk group and 50 of them defined as TD group.

3.2.2 The sentence repetition task (SRT)

In an attempt to measure students' performance in SRT, the two groups participated in SRT. In previous researches, SRT is commonly composed of 3 parts according to the syntactic complexity of the sentences: simple sentences (n= 5), featuring canonical subject-verb-object structures, complex sentences (n= 6), containing embedded clauses, such as object clauses or relative clauses, and passive sentences (n= 5), involving non-canonical word order [14,15]. The author followed the common approach of designing SRT and all the vocabulary in SRT in in the national high school English teaching curriculum.

3.3 Procedure

The experimental session took place one at a time in a quiet laboratory over a period of 30 min. Each participant gave informed consent and completed the ADC screening questionnaire. They then sat in front of a computer and brought on headphones. Instructions for the SRT were presented on the screen and verbally expounded by the experimenter. An instruction was given, instructing participants to carefully listen to each sentence and to orally produce it as precisely to possible immediately following a tone signal. Three practice trials with corrective feedback were offered to ensure understanding of the task. The 16 experimental sentences were then presented in a randomly varied order across participants. The session was audially recorded by a high-quality digital recorder for verbatim transcription and scoring.

3.4 Data scoring and analysis

Participant responding was transcribed and scored in a finely-grained method (0-2 points per item) in line with established SRT scoring protocols [7]. A point was award for correct and precise repetition. 1 was awarded for repetitions containing minor mistakes that maintained the basic syntactic frame and the meaning (for example, omission/substituted function words, tenses mistakes of verbs). 0 points were awarded for repetitions with syntactic structure changes, omissions/ substitution of content words, or any type of error whose changed the core meaning. Total scores (out of 32) and sub-scores for each type of sentence were calculated for each participant.

SPSS 29.0 was used to analyze the data. First and foremost, the descriptive analysis was performed to show the details of data. An independent sample t-test was then followed to reveal the differences in performance of English Sentence Repetition Performance between

the at-risk group and TD group. The exploratory factor analysis (EFA) analysis was conducted to find predictive factors of students' performance of SRT.

4 Results

4.1 The differences in SRT between at-risk group and TD group

To answer the first research question, the descriptive analysis of data was done and the description of ADC scores and SRT scores of at-risk and TD group is posted in Table 1 and Table 2.

Table 1. Descriptive analysis of at-risk group.

Names	Sample size	Minimum value	Maximum value	Mean	Standard deviation	Median
SRT score	50	3.000	16.000	10.760	3.217	11.000
Simple sentence score	50	3.000	10.000	7.240	1.506	7.000
Complex sentence score	50	0.000	7.000	3.140	1.980	3.000
Passive sentence score	50	0.000	2.000	0.380	0.567	0.000
ADC Writing	50	4.000	16.000	8.260	2.724	8.000
ADC Listening	50	5.000	16.000	8.420	2.851	8.000
ADC Reading	50	6.000	16.000	8.840	1.765	8.000
ADC Speaking	50	6.000	14.000	9.080	1.676	9.000

Table 2. Descriptive analysis of TD group.

Names	Sample size	Minimum value	Maximum value	Mean	Standard deviation	Median
SRT score	51	24.000	32.000	28.882	2.141	29.000
Simple sentence score	51	7.000	10.000	9.020	1.122	9.000
Complex sentence score	51	9.000	13.000	10.569	1.082	10.000
Passive sentence score	51	7.000	10.000	9.353	0.820	10.000
ADC Writing	51	0.000	4.000	1.471	1.286	1.000
ADC Listening	51	0.000	4.000	1.333	1.322	1.000
ADC Reading	51	0.000	5.000	2.608	1.457	3.000
ADC Speaking	51	0.000	3.000	2.059	1.287	3.000

From the descriptive analysis, it is revealed that the mean of the total SRT score of TD group is significantly higher than at-risk group, with the at-risk group mean of 10.760 and the TD group mean of 28.882, which is a significant difference. And then, the difference is

the smallest in Simple sentence tasks (7.240 and 9.020), is moderate in Complex sentence tasks (3.140 and 10.569), and is the biggest in Passive sentence tasks (0.380 and 9.353). There is no outliers in the two graphs, which proves the validity of the data.

Also, the independent sample t-test was conducted between the 50 students at-risk groupt and the 51 students TD group to reveal the difference between the at-risk group and TD-group. The result is shown in Table 3.

Table 3 The Independent Sample T-Test

Items	Groups	Mean	Standard deviation	mean difference	Difference 95% CI	t	df	p
SRT score	At-risk	10.76	3.22	-18.12	-19.206 ~ -17.039	-33.259	85.084	0.000**
	TD	28.88	2.14					
	Total	19.91	9.50					
Simple sentence score	At-risk	7.24	1.51	-1.78	-2.303 ~ -1.256	-6.743	99.000	0.000**
	TD	9.02	1.12					
	Total	8.14	1.59					
Complex sentence score	At-risk	3.14	1.98	-7.43	-8.063 ~ -6.795	-23.338	75.545	0.000**
	TD	10.57	1.08					
	Total	6.89	4.05					
Passive sentence score	At-risk	0.38	0.57	-8.97	-9.251 ~ -8.695	-64.034	89.070	0.000**
	TD	9.35	0.82					
	Total	4.91	4.56					

From Table 3, it can be seen that when using independent sample t-test to study the effect of Group on SRT score, Simple sentence score, Complex sentence score, Passive sentence score, differences in at risk group and TD group samples for SRT score, Simple sentence score, Complex sentence score, Passive sentence score all showed significant differences ($p<0.05$). The Group showed a significant 0.01 level of SRT score ($t=-33.259$, $p=0.000$), and specific comparative differences indicate that the average value of at-risk (10.76) is significantly lower than the average value of TD (28.88).

The at-risk group and TD group showed significant significance at the 0.01 level for the Simple sentence score ($t=-6.743$, $p=0.000$). The average value of at-risk (7.24) was significantly lower than that of TD (9.02).

The at-risk group and TD group showed significant significance at the 0.01 level for the Complex sentence score ($t=-23.338$, $p=0.000$). The average value of at-risk (3.14) was significantly lower than that of TD (10.57).

The at-risk group and TD group showed significant significance at the 0.01 level for the Pass sentence score ($t=-64.034$, $p=0.000$). The average value of at-risk (0.38) was significantly lower than that of TD (9.35).

In sum, the at-risk group and TD group shows significant difference in SRT performance, with TD group performing obviously better than at-risk group. The difference is the most obvious in Passive sentence score and the least obvious in Simple sentence score, with

Complex sentence in the middle, which means that the higher the syntactic complexity, the more significant differences of the performance the two groups have.

4.2 The predictive factors of SRT performance

In an attempt to answer the second research question, the author used the correlation analysis and regression analysis. The items in ADC scale can be divided into four factors: Speaking difficulty, listening difficulty, reading difficulty, and writing difficulty. In an attempt to answer the second research question, measuring their contribution to SRT performance, the author first used the correlation analysis and the result is shown in table 4.

Table 4. Pearson Correlation.

	SRT score
ADC Writing	-0.952**
ADC Listening	-0.950**
ADC Reading	-0.914**
ADC Speaking	-0.941**

* p<0.05 ** p<0.01

Correlation analysis was applied to uncover the correlation between SRT score and four items including ADC Speaking, ADC Listening, ADC Reading, and ADC Writing. The correlation coefficient between SRT score and ADC Speaking is -0.952, with significance at the 0.01 level. It means that there is a significant negative correlation between SRT score and ADC Speaking. Similarly, when it comes to SRT score and ADC Listening, SRT score and ADC Reading, SRT score and ADC Reading, the correlation coefficient between them turns out to be -0.950, -0.914 and -0.941, with 0.01 level significance, indicating a significant negative correlation, too.

And then, the regression analysis is also finished, which is shown in Table 5.

Table 5. The Regression Analysis.

The Regression Analysis (n=101)							
	Non-standard Coefficient		Standard Coefficient	t	p	Collinearity Diagnostics	
	B	Standard Deviation	Beta			VIF	Tolerance
Constant	32.636	0.420	-	77.627	0.000**	-	-
ADC Writing	-0.605	0.265	-0.255	-2.278	0.025*	25.102	0.040
ADC Listening	-0.645	0.248	-0.284	-2.599	0.011*	23.925	0.042
ADC Reading	-0.360	0.161	-0.133	-2.235	0.028*	7.104	0.141
ADC Speaking	-0.837	0.161	-0.337	-5.195	0.000**	8.416	0.119
R 2	0.952						

The Regression Analysis (n=101)							
	Non-standard Coefficient		Standard Coefficient	t	p	Collinearity Diagnostics	
	B	Standard Deviation	Beta			VIF	Tolerance
Adjust R 2	0.950						
F	F (4,96) =475.554, p=0.000						
D-W Value	1.825						

Note：Dependent variable = SRT score

* p<0.05 ** p<0.01

According to Table 5, ADC Writing, ADC Listening, ADC Reading, and ADC Speaking are independent variables, and SRT score is the dependent variable. The model formula for linear regression analysis is: SRT score=32.636-0.605*ADC Writing-0.645*ADC Listening-0.360*ADC Reading-0.837*ADC Speaking. The R-squared value of the model is 0.952, which means that ADC Writing, ADC Listening, ADC Reading, and ADC Speaking can explain 95.2% of the variation in SRT score. Also, the model passed the F-test (F=475.554, p=0.000<0.05). So that the model construction is meaningful.

The regression coefficient value of ADC Writing is -0.605 (t=-2.278, p=0.025<0.05), so that ADC Writing will have a negative impact on the SRT score in a moderate manner. As for ADC Listening, it is -0.645 (t=-2.599, p=0.011<0.05), which has a negative impact on the SRT score in terms of Moderation. The regression coefficient value of ADC Reading is -0.360 (t=-2.235, p=0.028<0.05), indicating that ADC Reading has a slight negative impact on SRT score. The regression coefficient value of ADC Speaking is -0.837 (t=-5.195, p=0.000<0.01), indicating that ADC Speaking has a significant negative impact on SRT score.

In summary, ADC Speaking, ADC Listening, ADC Reading, and ADC Writing all have a negative impact on the SRT score. The predictive effect of ADC Speaking is the most significant, followed by ADC Listening and Writing, while the predictive effect of ADC Reading is not significant.

5 Discussion

The findings of this research align with some previous research in that Chinese high school students at risk for developmental language disorder demonstrate significantly poorer SRT performance than their typically developing peers. The underlying reason is may lie in the systematic impact of developmental language disorder on second language acquisition processes, since it can lead to the consequences of shorter working memory, inability in controlling speech organs, difficulties in mental processing procedures and so forth, which also stands in Chinese contexts [4].

Different from former researches that mainly reports overall group differences, this paper takes a closer look at the details of the differences and finds out that the syntactic complexity can affect the differences. The performance gap widens as sentence structures become syntactically more complex. The plausible explanation is that syntactically complex sentences impose higher cognitive load [16]. Also, this paper explored the predictive factors of SRT performance according to the ADC scale and it turns out that the high risk of developmental speaking disorder serves as the strongest predictor of students' poor SRT

performance, and listening disorder as well as writing disorder have moderate predictability, which possibly because they share partial processing demands with SRT. But risk for reading disorder is not an important predictive factor. It is probably because speaking is a more comprehensive language activity, which involves ability in dealing with both input and output so that it is more similar to SRT [6]. While reading involves more preparation time and reaction time, which does not require such high language ability standard as SRT required. Especially in Chinese context, most of the time because of the high pressure of the college entrance exam, even a student who is highly vulnerable in English learning would be forced to memorize a large amount of vocabulary and meaning, so there might exist a mismatch between Chinese high school students' L2 reading ability and their real L2 ability. And writing and Listening are in the middle, which leads to SRT performance in varying degrees.

6 Conclusion

In conclusion, this study firstly supplements the previous studies of the differences in performance of SRT between students at risk for developmental language disorder and their TD peers by delving into Chinese high school context and thus extends the generalization of the former research finding. Secondarily, it takes a more detailed look at what are the differences and reveals that the syntactic complexity could positively affect the amount of the differences. In addition, this study explored the predictable factors of students' SRT performance and finds that the high risk for developmental speaking disorder is the most important predictive factor in SRT performance while reading is the least predictive, which can be ascribed to to specific characteristics of the Chinese high school system.

Limitations also exist in this study. The research context is Zhejiang province, where the economy as well as education is well- developed, so that it is not representative enough to stand for the whole case of China, since China covers a wider region, especially the north and west part where the education quality is far less developed. Also, the SRT is done only once and a one-time test might be too random and cannot fully represent students' real SRT performance. Future studies can try to cover the other parts of China to reach a more representative outcome and the tests are supposed to be repeated several times for a more reliable SRT performance result.

References

1. Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., Pickles, A., The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study, J. Child Psychol. Psychiatry, **57(11)**, 1247–1257 (2016)
2. Conti-Ramsden, G., Durkin, K., Toseeb, U., Postschool educational and employment experiences of young people with specific language impairment, Lang. Speech Hear. Serv. Sch., **49(2)**, 140–153 (2018)
3. Glasby, J., Graham, L. J., White, S. L., Tancredi, H., Do teachers know enough about the characteristics and educational impacts of Developmental Language Disorder (DLD) to successfully include students with DLD?, Teach. Teach. Educ., **119**, 103868 (2022)
4. Rinaldi, S., Caselli, M. C., Cofelice, V., D'Amico, S., De Cagno, A. G., Della Corte, G., ... & Zoccolotti, P. (2021). Efficacy of the treatment of developmental language disorder: A systematic review. Brain sciences, **11(3)**, 407.

5. Rujas, I., Mariscal, S., Murillo, E., Lázaro, M., Sentence repetition tasks to detect and prevent language difficulties: A scoping review, Children, **8(7)**, 578 (2021)
6. Leonard, L. B., Schroeder, M. L., The study of children with developmental language disorder beyond English: A tutorial, Lang. Acquis., **31(3–4)**, 180–198 (2024)
7. Polišenská, K., Chiat, S., Roy, P., Sentence repetition: What does the task measure?, Int. J. Lang. Commun. Disord., **50(1)**, 106–118 (2015)
8. Pratt, A. S., Peña, E. D., Bedore, L. M., Sentence repetition with bilinguals with and without DLD: Differential effects of memory, vocabulary, and exposure, Biling. Lang. Cogn., **24(2)**, 305–318 (2021)
9. Taha, J., Stojanovik, V., Pagnamenta, E., Sentence repetition as a clinical marker of developmental language disorder: Evidence from Arabic, J. Speech Lang. Hear. Res., **64(12)**, 4876–4899 (2021)
10. Scheidnes, M., Tuller, L., Using clausal embedding to identify language impairment in sequential bilinguals, Biling. Lang. Cogn., **22(5)**, 949–967 (2019)
11. Gao, J., Ma, S., Learning condition, linguistic complexity, and first language transfer in semi-artificial language learning: A conceptual replication and extension of Tagarelli et al. (2016), Stud. Second Lang. Acquis., **43(2)**, 355–378 (2021)
12. Kirby, A., Edwards, L., Sugden, D., Rosenblum, S., The development and standardization of the adult developmental co-ordination disorders/dyspraxia checklist (ADC), Res. Dev. Disabil., **31(1)**, 131–139 (2010)
13. Preacher, K. J., Rucker, D. D., MacCallum, R. C., Nicewander, W. A., Use of the extreme groups approach: A critical reexamination and new recommendations, Psychol. Methods, **10(2)**, 178 (2005)
14. Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., A comprehensive grammar of the English language, (Longman, Harlow, 1985)
15. Lecouvet, M., Non-canonical word order as a measure of syntactic complexity in advanced L2 German, Int. Rev. Appl. Linguist. Lang. Teach., **61(3)**, 877–907 (2023)
16. Rinaldi, S., Caselli, M. C., Cofelice, V., D'Amico, S., De Cagno, A. G., Della Corte, G., Zoccolotti, P., Efficacy of the treatment of developmental language disorder: A systematic review, Brain Sci., **11(3)**, 407 (2021)