Examining the Preliminary Validity of a Dynamic Assessment of Narrative Language in Mandarin Chinese

Lok Yee Sarah Cheung
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd
Part of the Communication Sciences and Disorders Commons

BYU ScholarsArchive Citation
Cheung, Lok Yee Sarah, "Examining the Preliminary Validity of a Dynamic Assessment of Narrative Language in Mandarin Chinese" (2022). Theses and Dissertations. 9608.
https://scholarsarchive.byu.edu/etd/9608

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.
Examining the Preliminary Validity of a Dynamic Assessment of Narrative Language in Mandarin Chinese

Lok Yee Sarah Cheung

A thesis submitted to the faculty of Brigham Young University in partial fulfillment of the requirements for the degree of Master of Science

Douglas Petersen, Chair
Shawn Nissen
Li Sheng
Connie Summers

Department of Communication Disorders Brigham Young University

Copyright © 2022 Lok Yee Sarah Cheung All Rights Reserved
ABSTRACT

Examining the Preliminary Validity of a Dynamic Assessment of Narrative Language in Mandarin Chinese

Lok Yee Sarah Cheung
Department of Communication Disorders, BYU
Master of Science

The purpose of this thesis was to examine the preliminary validity of a newly developed dynamic assessment of narrative language in Mandarin Chinese. Two studies are reported in this thesis. Study 1 included 31 second grade participants and Study 2 included 43 first grade participants. All participants were enrolled in a Chinese immersion program in an elementary school in Utah. A dynamic assessment of narrative language was administered to each participant in Mandarin Chinese. A teacher rating was also included in Study 1. Results indicated that the dynamic assessment investigated in this thesis demonstrated some similar characteristics with other valid dynamic assessments of narrative language. As hypothesized, participants in Study 1 made gains from pretest to posttest after the teaching phase. Gain score from pretest to posttest and static teacher rating did not significantly correlate with modifiability rating.

Modifiability rating and posttest score were significantly correlated in both studies. There is also no significant group difference between the participants in Study 1 and Study 2 on modifiability rating. These results are promising. However, more research will need to be conducted to further examine the dynamic assessment due to the limitations of the current studies.

Keywords: dynamic assessment, narrative language, oral language, school-age, Mandarin Chinese
ACKNOWLEDGMENTS

I would like to thank the following committee members for their guidance and contribution in writing this thesis: Dr. Petersen, Dr. Nissen, Dr. Sheng, and Dr. Summers. Their expertise and experience were vital to this thesis. I would also like to thank all the research assistants for their time and effort in collecting the data used in this project. Lastly, I would like to express my gratitude to my family and friends for their support throughout this project.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td>i</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>DESCRIPTION OF THESIS STRUCTURE AND CONTENT</td>
<td>viii</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Current Practices in Assessing Mandarin-Speaking Children With Language Disorders</td>
<td>1</td>
</tr>
<tr>
<td>Dynamic Assessment</td>
<td>2</td>
</tr>
<tr>
<td>Narrative Language</td>
<td>4</td>
</tr>
<tr>
<td>Dynamic Assessments of Narrative Language</td>
<td>5</td>
</tr>
<tr>
<td>Study 1</td>
<td>8</td>
</tr>
<tr>
<td>Method</td>
<td>9</td>
</tr>
<tr>
<td>Participants</td>
<td>9</td>
</tr>
<tr>
<td>Measures</td>
<td>9</td>
</tr>
<tr>
<td>Dynamic Assessment of Narrative Language</td>
<td>9</td>
</tr>
<tr>
<td>Pretest and Posttest</td>
<td>10</td>
</tr>
<tr>
<td>Teaching Phase</td>
<td>10</td>
</tr>
<tr>
<td>Modifiability</td>
<td>11</td>
</tr>
<tr>
<td>Teacher Rating</td>
<td>12</td>
</tr>
<tr>
<td>Test Administration: Fidelity</td>
<td>12</td>
</tr>
<tr>
<td>Study 2</td>
<td>12</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1  Demographic Information for Study 1 .................................................................9
Table 2  Study 1 Means, Range, and Standard Deviation of Pretest and Posttest Measures....15
Table 3  Study 1 Correlation Matrix of Dynamic Assessment Variables and
Teacher Rating......................................................................................................................16
DESCRIPTION OF THESIS STRUCTURE AND CONTENT

This thesis, *Examining the Preliminary Validity of a Dynamic Assessment of Narrative Language in Mandarin Chinese*, is written in a hybrid format. The initial pages of this thesis reflect requirements for submission to the university, and the remainder of the thesis report is presented as a journal article format. The annotated bibliography is included in Appendix A. Institutional Review Board (IRB) Approval is found in Appendix B. The story used in Study 1 is included in Appendix C. Appendix D provides information regarding scoring for the Dynamic Assessment of Narrative Language in Study 1. Pictures used for the teaching phase in Study 1 is found in Appendix E. Story grammar icons used for the teaching phase in Study 1 is found in Appendix F. The instructions and the script for the teaching phase used in Study 1 are found in Appendix G. The Modifiability rating form for Study 1 is included in Appendix H. Appendix I shows the teacher rating form for Study 1. Lastly, the story used for Study 2 is included in Appendix J.
Introduction

Mandarin Chinese is one of the most widely used languages in the world with over 1 billion speakers in China, Taiwan, Singapore, Hong Kong, and other Eastern and Western countries. With over 900 million native speakers and 200 million non-native speakers, many of the population are children who are currently attending schools. According to the National Population Census conducted by the People's Republic of China in 2021, there are over 250 million children between the age of 0 to 14 (National Bureau of Statistics of China, 2021). Currently, there is no national census data available on the prevalence of children with language disorders in China (Wu et al., 2020). However, given a 7% prevalence rate of developmental language disorder (DLD) in English-speaking children (Tomblin et al., 1997) and assuming DLD is a worldwide phenomenon that is equally prevalent across languages and cultures (Armon-Lotem et al., 2015), over 10 million 0 to 14-year-old children in China are estimated to have DLD (National Bureau of Statistics of China, 2021). The current number of Mandarin-speaking children who are estimated to have language disorder demonstrate a need for speech-language pathologists and educators to validly assess and identify this population.

Current Practices in Assessing Mandarin-Speaking Children With Language Disorders

Standardized norm-referenced assessments are commonly used for identifying children with language disorder (Caesar & Kohler, 2007). The profession of speech-language pathology and research on language disorder in Mandarin are fairly new in China. Some Mandarin standardized norm-referenced language assessments currently being used include the Preschool Language Scale for Developmentally Delayed Children (Lin et al., 2008), The Diagnostic Receptive and Expressive Assessment of Mandarin (DREAM) (Liu et al., 2017), and The Mandarin Clinical Evaluation of Language for Preschooler’s Core Scale (MCELP-CS; Wu et al.,
According to studies done with English-speaking children, most norm-referenced assessments of language have poor sensitivity and specificity (Spaulding et al., 2006). Sensitivity is a measure of how well an assessment can accurately identify children with language disorders and specificity is a measure of how well an assessment can correctly identify children without language disorders. The sensitivity and specificity of a diagnostic assessment should be at or above 80% in order to be considered acceptable (Spaulding et al., 2006).

In addition to having adequate sensitivity and specificity, norm-referenced tests should also include a normative sample that represents the intended population. Most Mandarin norm-referenced assessments only include children who are monolingual Mandarin-speakers from the majority Han culture in the normative sample. Even though Mandarin is the official language, there are many different Chinese dialects spoken in China such as Cantonese, Hakka, Shanghainese, and Hainanese. Many people learn these dialects from home as their first language and learn Mandarin from school as a second language. Moreover, there are 55 ethnic minorities in China with cultures that are different from the mainstream Han culture. The current norm-referenced assessments could be linguistically or culturally biased against children who are Mandarin language learners or children who are from an ethnic minority group. There is a need for a more linguistically and culturally appropriate language assessment that can be used to identify Chinese-speaking children with language disorders.

**Dynamic Assessment**

Instead of assessing a child’s current knowledge and ability through a norm-referenced assessment, dynamic assessment measures a child’s learning ability. Dynamic assessment often follows a pretest, teaching, and then posttest phase, where students are provided brief intervention and their response to that intervention is measured. Dynamic assessment is based on
the principles of the Zone of Proximal Development (Vygotsky, 1978) and the Mediated Learning Experience (Feuerstein, 1979; Ukrainetz et al., 2000). A child’s zone of proximal development falls between what a child can do independently and what a child can do with appropriate assistance. Vygotsky’s theory indicates that learning occurs when scaffolding is provided when a more capable person assists a child in the zone of proximal development as necessary and decreases the level of support when the child becomes more competent. When this theory is applied to dynamic assessment, it suggests that children should make gains from pretest to posttest as a result of scaffolding and direct instruction with focused questions and positive interactions within their zone of proximal development. Feuerstein’s Mediated Learning Experiences (MLE) is fundamental to the teaching phase of dynamic assessments. MLE involves intentional instruction focusing on facilitating independent learning in children by monitoring their responses to intervention. In contrast to Vygotsky’s focus on gains and changes in tasks, Feuerstein’s approach emphasizes the child’s behavior and response to instructions during the mediated learning experience. When applied to dynamic assessment, a child’s response to prompts, awareness of errors, disruptive behavior, level of frustration, attention, transfer of strategies within and between tasks, and motivation can be observed during the teaching phase to determine a child’s modifiability which is a measure of examiner effort and child responsivity (Peña et al., 2006; Peña et al., 2014; Petersen et al., 2017). It has been theorized that because modifiability is a measure of a child’s learning potential, factors such as age, race, ethnicity, gender, socioeconomic status should not affect the modifiability outcome (Feuerstein, 1979).

Both Vygotsky’s and Feuerstein’s approaches are taken into account in dynamic assessments by using a test-teach-retest model. During the pretest, the examiner is able to determine how well a child can perform a task independently. Throughout the teaching phase,
the examiner provides support from high level to low level gradually by reducing the amount of
assistance provided. During this process, the examiner monitors how the child responds to
instructions to determine modifiability. Learning is indicated when the child applies skills taught
by the examiner when given less assistance. However, when the child demonstrates difficulty
producing behaviors taught by the examiner despite maximum assistance, it suggests that the
child experiences difficulty with learning. The posttest is administered to provide an indicator of
gain from the pretest and to provide information on how much the child learned from the
teaching phase.

Instead of comparing a child’s performance to their peers in norm-referenced
assessments, dynamic assessment measures learning ability and modifiability which can reduce
bias against culturally and linguistically diverse children. Dynamic assessment can also avoid
confounding variables such as socioeconomic status and poorly representative normative
samples. Since dynamic assessment has potential benefits for all children, especially culturally
and linguistically diverse children, studies have been conducted to investigate the potential of
using dynamic assessments to identify language disorder in children (Kapantzoglou et al., 2012;

**Narrative Language**

Narrative language has been a major focus of dynamic assessment of language research
because it is functional and also includes complex grammatical forms and vocabularies (Colozzo
et al., 2011; Squires et al., 2013). Narratives involve the use of story grammar elements, which
are integral components of a story. Story grammar elements include character, setting, problem,
attempt, feeling, consequence, and resolution. Narratives also involve the use of academic
language such as adverbial and adjectival subordinate clauses to provide context and details for a
story. Since narratives are linguistically complex, difficulty producing and understanding narratives is an indicator of language disorder. A study on narrative abilities of Mandarin-speaking children with language disorder (Torng & Sah., 2020) indicated that children with language disorder demonstrated significantly fewer story grammar components, evaluative comments, and cohesion than children with typically developing language. Sheng et al. (2020) also assessed the narrative skills of Mandarin-speaking children who were at risk (AR) for language disorder. Results indicated that children in the AR group demonstrated significantly lower overall story structure scores than children with typically developing language. Lexical diversity and sentence complexity were also noted to be able to differentiate between children in the AR group and children with typically developing language.

**Dynamic Assessments of Narrative Language**

Several studies have provided evidence supporting the use of English dynamic assessments of narrative language, especially with culturally and linguistically diverse children (e.g., Hasson & Joffe, 2007; Patterson et al., 2013; Peña & Iglesias, 1992, Pena et al., 2006; Pena., 2014; Petersen et al., 2017; Ukrainetz et al., 2000). These studies suggest that dynamic assessment of narrative language can accurately identify language disorder in children from culturally and linguistically diverse backgrounds. Instead of measuring fractionalized language skills, narratives measure integrates language in a functional communicative context which a child uses on a daily basis (Ukrainetz et al., 2000; Westby, 1985).

A recent meta-analysis (Orellana et al., 2019) reported that English dynamic assessments of narrative language consistently demonstrate high sensitivity and specificity with bilingual students who are from culturally and linguistically diverse backgrounds. Findings are as follows: Kramer et al. (2009): Sensitivity- 100%, Specificity- 91.7%; Pena et al. (2009): Sensitivity-
88.9%, Specificity- 88.9%; Petersen et al. (2017): Sensitivity- 100%, Specificity- 94%. Orellana et al. (2019) also investigated which components of a dynamic assessment can differentiate diverse children who have language disorder from those with typically developing language. The results indicated that children with typically developing language received higher scores on some dynamic assessment variables than children with language disorder. However, both children with and without language disorder tend to make gains from pretest to posttest with no significant group differences on gain scores. Thus, gains from pretest to posttest have not been able to accurately differentiate children with or without language disorder (Peña et al., 2006; Peña et al., 2014; Ukrainetz et al., 2000).

Research has also indicated that children with typically developing language receive significantly higher modifiability ratings than children with language disorder. Children with typically developing language have mean high scores on modifiability ratings with a relatively small standard deviation, while children with language disorder usually receive significantly lower modifiability ratings. For example, a study by Petersen et al. (2017) found that children with typically developing language had a mean total modifiability index of 12.84 in a 14-point scale with a standard deviation of 1.64, while children with language disorder received a mean total modifiability index of 5.20 and a standard deviation of 3.49. A recent study by DeRobles (2021) also found that children with typically developing language received a mean modifiability score of 3.77 in a 5-point scale with a standard deviation of 0.43, while children with language disorder were given a mean modifiability score of 2.28 with a standard deviation of 0.99. These studies indicated that modifiability can be used to differentiate children with or without language disorder. They also reported that gain scores from pretest to posttest have not been significantly
correlated with modifiability ratings. Pretest scores and other static assessments such as teacher ratings also do not correlate with modifiability ratings.

Petersen et al. (2017) investigated the diagnostic accuracy of an English dynamic assessment of narrative with bilingual Spanish-English speakers and sought to identify the combination of scores (modifiability rating, gain score, posttest score, teaching duration) that yielded the highest classification accuracy of language disorder. Similar to previous studies (Peña et al., 2006; Peña et al., 2014), results from Petersen et al. (2017) found that the overall modifiability rating was the best classifier with 100% sensitivity and 100% specificity after two, 20-minute dynamic assessment sessions and 100% sensitivity and 88% specificity after one 20-minute dynamic assessment session. The results of the study also indicated that both sensitivity and specificity were over 90% when combining any two scores from the three predictor variables (posttest scores, modifiability ratings, or teaching duration) after just one 20-minute session.

Prospective studies on dynamic assessment of narrative language should reflect the key characteristics of valid dynamic assessments. Namely, both children with and without language disorder make pretest to posttest gains, children with typically developing language have mean high scores on modifiability ratings with a relatively small standard deviation, gain scores from pretest to posttest do not correlate significantly with modifiability ratings, and pretest scores and other static tests such as teacher ratings do not correlate with modifiability. Future studies should particularly identify which dynamic assessment variables are most predictive of language ability, use a one-gate design where researchers are blind to language ability prior to testing, and cross-validate and replicate previous findings (Orellana et al., 2019). The long-term goal of this project is to develop and validate a dynamic assessment of narrative language in Mandarin Chinese that has limited cultural and linguistic bias for children who are bilingual in Mandarin Chinese and
Chinese dialects in China and in English-speaking countries. This is the first phase of the project. The purpose of the following two studies is to examine if the Mandarin Chinese adaptation of a dynamic assessment of narrative language has similar characteristics as other valid dynamic assessments.

Study 1

We hypothesized that the Mandarin Chinese dynamic assessment of narrative language would have similar characteristics as the dynamic assessment which it was adapted from (Petersen & Spencer, 2014) and other valid dynamic assessments. The following research questions were investigated:

1. Do typically developing second-grade Mandarin language learners make gains from pretest to posttest in the dynamic assessment of narrative language?

2. Do these typically developing students have a high mean score on modifiability rating with a relatively small standard deviation?

3. Do gains from the pretest to posttest, static teacher rating of current language ability, and posttest score correlate with modifiability rating?

Based on the results of Petersen et al. (2017), which this study was adapted from, and other studies on valid dynamic assessments of narrative language, we hypothesized that the typically developing second-grade students would make gains from pretest to posttest in the dynamic assessment of narrative language, with a high mean score on modifiability rating and a relatively small standard deviation. We also hypothesized that gain score and static teacher rating would not correlate with modifiability rating. However, we hypothesized that posttest score would correlate with modifiability rating.
Method

Participants

This study included 31 second-grade students from Chinese immersion classrooms in an elementary school in Utah. Participants' characteristics were displayed in Table 1. Twenty-one were male and ten were female with an average age of 7.3 years old. None of the participants had an individualized education program (IEP) in any area of classification at the time of the study. English was the native language for 27 participants, and 4 participants spoke other languages at home. Most participants began learning Mandarin Chinese when they entered first grade which was approximately one year prior to the study.

Table 1

Demographic Information for Study 1

<table>
<thead>
<tr>
<th>Gender</th>
<th>n=31 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>21 (68%)</td>
</tr>
<tr>
<td>Female</td>
<td>10 (32%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Home Language</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>27 (87%)</td>
</tr>
<tr>
<td>Mandarin</td>
<td>3 (10%)</td>
</tr>
<tr>
<td>Korean</td>
<td>1 (3%)</td>
</tr>
</tbody>
</table>

Measures

Dynamic Assessment of Narrative Language

All participants were assessed using a Mandarin Chinese adaptation of the language subtest of the Petersen & Spencer (2014) Predictive Early Assessment of Reading and Language (PEARL) dynamic assessment. A team of four trained, Mandarin-speaking research assistants administered the dynamic assessment in quiet rooms in the participants' school. The language
subtest included four steps: a pretest, a teaching phase, a set of modifiability rating scales, and a posttest. All scoring of the assessment was done in real time. All testing was completed in one day. For most participants, the entire testing required approximately 8 minutes. However, the length varied depending on the responsiveness of each participant, ranging from 6 minutes to 16 minutes.

**Pretest and Posttest**

The pretest and posttest involved the examiner reading a brief story and having the participants retell the narrative. The participants were assessed on their inclusion of story grammar elements (i.e., character, setting, problem, feeling, plan, action, consequence, ending, and end feeling), episode complexity (e.g., problem, attempt, consequence), and elements of language complexity (e.g., subordinate clauses). The pretest narrative retell was scored in real-time using a point system, with a maximum score of 25 points. This maximum score was comprised of the story grammar subtotal score, the episode subtotal score, and the language complexity scores. The posttest followed the same procedure as the pretest with the same story. No examiner support was given during the pretest or posttest.

**Teaching Phase**

The teaching phase included four steps which were designed to help participants to learn to produce a complete narrative episode that involves a problem, an attempt, a consequence, and an ending independently and to improve their language complexity. In the first step, the examiner retold the story used in the pretest while simultaneously pointing to a set of corresponding pictures and story grammar icons. The examiner pointed to each icon and modelled the story using the story grammar elements explicitly (e.g., In Mandarin: “This is Tony. He is the character in this story.”). Following this part of the instruction, the participant
practiced retelling the story using the same pictures and icons. The examiner provided support to the participant to include all story grammar elements and language complexity targets. Once the participant completed the retell with pictures and icons, they moved on to the next step of the teaching phase, where the pictures were removed. Then, the participant was asked to retell the story again using only the icons. The examiner continued to provide support and prompts when needed. During the last step of the teaching phase, the participant was asked to retell the story without the pictures or icons. The examiner provided support when necessary.

During any step of the teaching phase, if a participant omitted or skipped a story grammar element, the examiner immediately stopped the participant from retelling the story and provided a level 1 prompt, which was an open-ended question (e.g., In Mandarin: “Where is Tony?”). If the participant did not respond to the level 1 prompt, the examiner would, provide a level 2 prompt which entailed modelling an appropriate response and having the participant repeat it (e.g., In Mandarin: “Tony was at the park. Repeat after me, Tony was at the park.”). Following either prompt, the examiner instructed the participant to go back one step in the story (back one story grammar element) and start telling the story from that point while including the missing story grammar element. In addition to focusing on story grammar elements, the examiner also encouraged the participant to increase language complexity by prompting the use of subordinating conjunctions such as because, when, or after (e.g., “Why was Tony crying? Because he fell from his bike, and it hurts.”).

**Modifiability**

After the teaching phase, the examiner rated the participant’s modifiability, which quantified their ability to learn language using a 5-point modifiability rating scale. The examiner rated the student on the following criteria: response to prompts, confidence, disruptions, and rate.
Then, the examiner rated the student’s overall responsiveness during the teaching phase on a scale of 0-4 which reflected the final modifiability score. A score of 4 represented relative ease in learning while a 0 represented difficulty learning.

**Teacher Rating**

A teacher rating questionnaire was sent to the participants’ teacher prior to the administration of the dynamic assessment. The teacher was asked to rate each participant’s current ability to understand or comprehend Chinese.

**Test Administration: Fidelity**

A team of research assistants who were fluent in Mandarin in the communication disorders program at Brigham Young University were trained to administer and score the dynamic assessment. They received extensive training over approximately 5 hours. The training included practice sessions administering and scoring the assessments on fellow research assistants playing the role of children. Each research assistant was required to demonstrate competence to conduct the testing procedures and scoring with 95% accuracy for the dynamic assessment. A speech-language pathologist with extensive experience in the development and validation of dynamic assessments observed and provided assistance throughout the training process.

**Study 2**

The purpose of study 2 was to examine a simplified version of the Mandarin Chinese dynamic assessment of narrative language used in Study 1 for even younger children or for children with less exposure to Mandarin. We hypothesized that the simplified version would share similar characteristics with the assessment used in Study 1 and with other valid dynamic assessments. The following research questions were investigated:
1. Do typically developing first-grade Mandarin language learners have a high mean score on modifiability rating with a relatively small standard deviation?

2. Do posttest score correlate with modifiability rating?

3. Is there a significant difference between first and second-grade participants on modifiability rating?

Similar to Study 1, we hypothesized that the typically developing first-grade students would demonstrate a high mean score on modifiability rating with a relatively small standard deviation. Posttest score would correlate with modifiability rating. We also hypothesized that there would not be a significant difference between the first and second-grade participants on modifiability rating because modifiability should not be affected by any demographic information.

**Method**

**Participants**

This study included 43 first-grade students from Chinese immersion classrooms in an elementary school in Utah. Twenty-five were male and eighteen were female with an average age of 6;4 years old. Similar to Study 1, none of the participants had an individualized education program (IEP) in any area of classification at the time of the study. English was the native language of all participants. Most participants began learning Mandarin Chinese approximately 3 months before they were being assessed.

**Measure**

Since the students were still at the beginning stage of learning Chinese, a simplified Mandarin Chinese dynamic assessment patterned after the Petersen & Spencer (2014) Predictive
Early Assessment of Reading and Language (PEARL) dynamic assessment was administered to all the participants.

The simplified version followed a similar pattern used in Study 1 and included three steps: a teaching phase, a set of modifiability rating scales, and a posttest. A pretest was not used in this study due to the participants still being at the beginning stages of learning Mandarin and would not be able to fully participate in the pretest. The assessment used in Study 2 also had the same scoring system as Study 1. The narrative used was a simplified version of the narrative used in Study 1 with fewer adjectives, adverbs, and subordinating conjunctions. All administration was conducted in Mandarin Chinese.

**Test Administration: Fidelity**

The team of research assistants who received extensive training and collected data for Study 1 participated in data collection for Study 2. Therefore, no other training was provided for Study 2. However, a speech-language pathologist with extensive experience in dynamic assessments observed and provided assistance throughout the data collection process.

**Results**

**Study 1 Data Analysis**

Data were analyzed using the Statistical Package for Social Sciences (SPSS version 27.0; IBM Corp., 2020). We examined if the typically developing second grade students made gains from pretest to posttest in the dynamic assessments of narrative language (Question 1). For the pretest, the participants had a mean score of 4.93 with a range of 0-19 and a standard deviation of 5.71. For the posttest, the mean score was 13.84 with a range of 3-21 and a standard deviation of 5.03 (Table 2). These results indicated that the participants made gains in the posttest that were nearly three times the score they received in the pretest of the dynamic assessment.
### Table 2

*Study 1 Means, Range, and Standard Deviations of Pretest and Posttest Measures*

<table>
<thead>
<tr>
<th></th>
<th>Mean Score</th>
<th>Range</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretest (n=31)</strong></td>
<td>M = 4.93</td>
<td>0-19</td>
<td>5.71</td>
</tr>
<tr>
<td><strong>Posttest (n=31)</strong></td>
<td>M = 13.84</td>
<td>3-21</td>
<td>5.03</td>
</tr>
</tbody>
</table>

*Note.* Maximum score= 25.

We also examined if the typically developing participants had a high mean score on the modifiability rating with a relatively small standard deviation (Question 2). Results indicated that the participants had a mean modifiability rating of 2.82 with a standard deviation of 0.83 and a range of 1-4.

A Pearson correlation coefficient analysis was conducted to determine the correlation among the modifiability rating, gain from the pretest to posttest, the static teacher rating, and the posttest score (Question 3, Table 4). The results indicated that the posttest score was significantly correlated with the modifiability rating ($r = .598, p = .01$). The results also showed that the modifiability rating was not significantly correlated with the gain score from pretest to posttest ($r = .117$) or teacher rating ($r = .243$). The gain score from pretest to posttest also did not correlate significantly with teacher rating ($r = -.158$), or posttest score ($r = .353$). The teacher rating did not correlate significantly with the posttest score ($r = .276$).
Table 3

Study 1 Correlation Matrix of Dynamic Assessment Variables and Teaching Rating

<table>
<thead>
<tr>
<th></th>
<th>Gain Score</th>
<th>Modifiability Rating</th>
<th>Teacher Rating</th>
<th>Posttest Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain Score</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modifiability Rating</td>
<td>.117</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher Rating</td>
<td>-.158</td>
<td>.243</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Posttest Score</td>
<td>.353</td>
<td>.598**</td>
<td>.276</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* Gain score = gain score from pretest to posttest.

Study 2 Data Analysis

Results indicated that typically developing first-grade students in study 2 had a mean modifiability rating of 2.38 with a standard deviation of 1.04 and a range of 0-4 (Question 1). To determine the correlation between modifiability rating and posttest scores (Question 2), a Pearson correlation coefficient analysis was conducted. The result indicated that modifiability rating was significantly correlated with posttest score ($r = .848, p = .01$).

An independent samples $t$-test was also conducted to examine whether there was a significant difference between the modifiability ratings of the first and second grade students (Question 3). The results showed that there was no significant difference between the modifiability ratings of first grade and second grade students ($t = -1.943, F(72,1) = 3.26, p = .056$).

Discussion

Study 1 and Study 2 served as initial explorations into adapting and developing a dynamic assessment of narrative language in Mandarin Chinese that shares similar characteristics with other valid dynamic assessments of narrative language. Results from both
Study 1 and Study 2 indicated that there were some similarities between the Mandarin Chinese dynamic assessment and other valid dynamic assessments.

As hypothesized, second grade students from Study 1 made gains from pretest to posttest after the teaching phase. These findings were aligned with previous studies. For example, Peña et al. (2006) found that both children in the typical development group and language impairment group in their study made gains from pretest to posttest in a dynamic assessment of narrative language. These findings provide additional support for Vygotsky’s ZPD, which postulates that children will make gains from pretest to posttest with adult support. It appears that even children with language disorder will make significant gains from pretest to posttest in a dynamic assessment and that the gains that they make will not be significantly less than the gains made by children with typically developing language. This is presumably due to the intensive individualized instruction provided to the children during the teaching phase. Although Study 1 and Study 2 only included typically developing children, the gains noted from pretest to posttest for our sample are promising.

We also found that the modifiability rating and posttest score were significantly correlated in both studies. Results indicated however that gain score from pretest to posttest and static teacher rating did not significantly correlate with modifiability ratings. These findings have been found in previous studies on dynamic assessments of narrative language. A study by Henderson et al. (2018) and a meta-analysis by Orellana et al. (2019) showed that modifiability rating was significantly correlated with posttest score across several dynamic assessment studies. These findings support Fuerstein’s postulation that learning potential can best be revealed through the MLE (teaching phase) of a dynamic assessment. The modifiability rating appears not to be an indicator of current performance, nor it is an indicator of gains from pretest to posttest.
Instead, modifiability is an indicator of how difficult it was for the children to make those gains which reveal language disorder.

In the current study, when comparing the modifiability rating between the first and second grade students, it was noted that the difference was not significant. This result was also expected because modifiability is a measure of a child’s learning potential which should not be affected by any demographic information such as age. This finding suggests that the same dynamic assessment and its accompanying modifiability rating scale may possibly be useful across multiple grades. It is possible however that younger children will receive lower modifiability ratings due to developmentally appropriate behaviors. Future research should investigate this potential variation in modifiability across ages.

However, the mean modifiability ratings in both Study 1 and Study 2 for our typically developing participants were lower than anticipated. The modifiability standard deviation in both studies was also larger than expected. Previous studies such as Petersen et al. (2017) and DeRobles (2021) found that children with typically developing language demonstrated a higher mean modifiability score and smaller standard deviation than children with language disorder. Because we did not included children with language disorder in our study, we do not know what their mean modifiability score would have been for them. However, if the mean modifiability score for the typically developing children is low, then there is a higher possibility that the children with language disorders will receive modifiability scores that overlap with the typically developing children, leading to poor sensitivity and specificity. The results on modifiability ratings from this current study could have been influenced by the participants’ limited Mandarin at the time of the assessment. This suggests that the examiners were not exclusively focused on whether the participants had a language disorder or not during the teaching phase but were
adversely influenced by the participants’ limited Mandarin language ability. It may be necessary
to further clarify that the modifiability rating should be unrelated to static language ability.

**Clinical Implications**

**Current Need for More Valid Assessments**

Since speech-language pathology is still a developing field in China, Mandarin language
assessments that are currently available demonstrate limitations that could be biased towards
children from diverse cultural and linguistic backgrounds. There is a need for more valid
assessments that can accurately identify Mandarin-speaking children with language disorders.
Previous studies (e.g., Hasson & Joffe, 2007; Patterson et al., 2013; Peña & Iglesias, 1992; Pena
et al., 2006; Pena et al., 2014; Petersen et al., 2017; Ukrainetz et al., 2000) have shown
promising results of utilizing dynamic assessment of narrative language to accurately identify
language disorder in children from culturally and linguistically diverse backgrounds. The results
from this study also indicated that the dynamic assessment of narrative language in Mandarin
Chinese shared, to some extent, similar characteristics with valid dynamic assessments of
narrative language used in previous studies.

The dynamic assessment used in this study was an efficient assessment that took
approximately 15 minutes to administer. Instead of assessing what a student currently knows
using static assessments, which could be affected by factors such as language difference, cultural
difference, socioeconomic status, and parental level of education, Chinese speech-language
pathologists may be able to utilize dynamic assessment to assess a student’s ability to learn
language.
The Role of Dynamic Assessment in Informing Language Treatment

Dynamic assessments can provide clinically relevant information for intervention. During the teaching phase, speech-language pathologists are able to assess students’ responsiveness during intervention and identify any breakdowns that the students experience. During this process, clinicians are able to determine a student's zone of proximal development, which provides information on how much support is needed for the student to be successful. Results of a dynamic assessment focusing on narrative language could also provide information that could help generate functional language goals. Oral narrative language includes academic language features that students would need to be competent in to be successful in school (Petersen et al., 2017).

Study Limitations

Overall, results from both Study 1 and Study 2 indicated that there were similarities between the dynamic assessment of narrative language in Mandarin Chinese and other valid dynamic assessment of narrative language in previous studies. However, there are limitations that may have affected the results of the studies. There were only 74 participants in the current studies. Most of the participants were beginning Chinese language learners who were enrolled in a Chinese immersion program in an elementary school at the time of data collection. The majority of our participants did not have the language environment similar to children who live in China or children who are from Chinese-English bilingual families in English-speaking countries to get exposure of Mandarin outside of school. Moreover, none of our participants had an IEP for language disorder at the time of the studies. Therefore, we were not able to determine the sensitivity and specificity of the dynamic assessment. Furthermore, reliability of the dynamic assessment was not analyzed.
**Future Research**

This initial investigation of adapting and developing a dynamic assessment of narrative language in Mandarin Chinese was part of a larger research study. The long-term goal of this project is to develop a dynamic assessment that has limited cultural and linguistic bias for children who are bilingual in Mandarin Chinese and Chinese dialects in China and also for children who are bilingual in Mandarin Chinese and English in English-speaking countries. The results of both Studies 1 and 2 indicated that the dynamic assessment of narrative language in Mandarin Chinese shared some similar characteristics with other valid dynamic assessments of narrative language. The results are promising. However, due to the limitations of this study, the results cannot be applied to all Mandarin-speaking children.

Future research will aim to improve the dynamic assessment by including questions for the examiners to answer about whether it is a language disorder or a language difference that is impacting modifiability ratings. Future study will also include a larger sample size with children who are bilingual in Mandarin Chinese and Chinese dialects in China and with children who are bilingual in Mandarin Chinese and English in English-speaking countries. We also plan to recruit children with and without language disorders to analyze the sensitivity and specificity of the dynamic assessment.
References


APPENDIX A

Annotated Bibliography


**Objective:** The purpose of this study was to identify the linguistic deficits in Mandarin-speaking children with language impairment using narrative analysis.

**Method:** This study included 18 Mandarin-speaking children with language impairment between the age of 4;03 and 7;11 and 18 Mandarin-speaking children with typical language between the age of 4;03 and 7;09. Each child was administered the experimental version of the Mandarin Expressive Narrative Test (MENT). During the assessment, the examiner presented pictures one by one in sequence to each child. After looking at all the pictures, each child was asked to tell the story with pictorial support. A total of 3 stories were included in the assessment. All narrative samples were video-recorded and transcribed. The samples were coded using the Computerized Language Analysis software for seven macrostructure elements (character, setting, initiating event, internal response, plan, action series, and consequence), general microstructure (TNU, TNW, NDW, MLU, usage of complex sentences, and grammaticality), and Mandarin-specific fine-grained microstructure (classifiers, perfective and progressive aspect markers, negation, “ba” structure, and “bei” structure).

**Results:** For macrostructure, there was a significant group difference between the LI group and the TD group on the macrostructure total scores. Compared to children in
the TD group, children in LI group demonstrated weakness in five macrostructure elements including character, setting, internal responses, action series, and consequence. For microstructure, children in the LI group demonstrated shorter MLU and produced fewer NDW with large effect sizes. However, there were no significant differences between the groups on TNW and TNU. For Mandarin-specific fine-grained microstructure, children in LI group demonstrated lower production of perfective aspect markers with large effect size. They also demonstrated inaccurate usage of the Mandarin passive “bei” structure and very low usage of classifiers.

*Relevance to current work:* The results from this study provide information on some difficulties that Mandarin-speaking children with LI may experience during a narrative assessment.


*Objective:* This purpose of this study was to introduce dynamic assessment into the field of speech-language pathology and propose it as an alternative means of language assessment.

*Method:* This study used a test-teach-retest approach with modifiability scores (dynamic assessment) for vocabulary testing with Puerto Rican and African American preschool children to identify children with and without language disorders.

*Results:* The result indicated that there were no significant differences on the pretest scores. However, the pre-posttest score pattern and modifiability scores yield promising results regarding the classification accuracy of children with and without language disorders.
Relevance to current work: The results from this study indicated that dynamic assessment is a promising alternative to current static measures in identifying children with language disorders, especially children from diverse cultural and language backgrounds.


Objectives: The purpose of this meta-analysis was to examine the current use of dynamic assessment and its diagnostic accuracy for diagnosing bilingual children with language disorders.

Method: The diagnostic accuracy of 7 studies on dynamic assessment was analyzed. The studies included bilingual children from 3 to 8 years old. Different areas of language such as narratives, labeling single words, ability to learn new words, and morpheme rule learning were targeted in the 7 studies.

Results: The results showed that participants with typically developing language performed better on the dynamic assessments used in the studies than participants with language disorders. Participants with typically developing language also showed significantly higher modifiability ratings than participants with language disorders. However, gain scores from pretest to posttest were not able to differentiate between children with or without language disorders. All seven studies showed high sensitivity and specificity.

Relevance to current work: The results from this meta-analysis indicated that there was evidence to support the use of dynamic assessment to diagnose bilingual
children with language disorders. Posttest scores and modifiability were able to
differentiate children with and without language disorders.

Patterson, J., RodríGuez, B., & Dale, P. (2013). Response to dynamic language tasks among
typically developing Latino preschool children with bilingual experience. *American
Journal of Speech-Language Pathology, 22*(1), 103–112. https://doi.org/10.1044/1058-
0360(2012/11-0129)

*Objectives:* The purpose of this study was to examine if typically developing bilingual
preschool children would show evidence of learning during the teaching phase of a
dynamic assessment of language.

*Method:* This study included 32 typically developing, bilingual 4-year-old children. 16 participants were Spanish-dominant and the other 16 participants were English-dominant. All participants participated in a dynamic assessment with tasks on
novel word learning, semantic, and phonological awareness. The dynamic assessment
was administered in the language that the participant had the most exposure to.

*Results:* The results showed that the participants performed significantly higher
on the final test items than the initial test items on the novel word learning and semantic
tasks. However, there was no significant difference between the initial test items and final
test items on the phonological awareness task. These findings indicated that the increase
in performance from graduated promptings during the teaching phase could be an
indicator of modifiability.

*Relevance to current work:* This study indicated that typically developing children
made gains from pretest to posttest in a dynamic assessment. Graduated promptings
which were used during the teaching phase of the dynamic assessment facilitated the
gains which suggested modifiability.

Peña, E., Gillam, R., & Bedore, L. (2014). Dynamic assessment of narrative ability in English
accurately identifies language impairment in English language learners. *Journal of
Speech, Language, and Hearing Research, 57*(6), 2208–2220.
https://doi.org/10.1044/2014_jslhr-l-13-0151

**Objective:** The purpose of this study was to assess the classification accuracy of a
dynamic assessment of narrative language for school-age students who were English
language learners.

**Method:** This study included 54 children. 18 were Spanish-English-speaking with
language impairment, 18 were typically developing controls matched by age, gender,
IQ, and language experience, and 18 were in a comparison group who were matched for
age and language experience. Each child was administered the dynamic assessment of
narrative language in English. The dynamic assessment used a pretest-teach-retest model
with two teaching phases. Macrostructure elements (story components, story ideas and
language, and episode structure) and microstructure elements (TNW, NDW, MLU,
percentage of ungrammatical utterances, and number of main verbs) were analyzed.

**Results:** The results indicated that children in the LI group had significantly lower
scores than children in the NL-Match group and children in the NL-Compare group.
Children in the LI group also had a higher (required more support during the teaching
phases) modifiability score than children in the other two groups. For classification
accuracy, discriminant analyses indicated that a combination of the DA story scores,
modifiability rating, and ungrammaticality from a language sample yielded 80.6% to 97.2% classification accuracy.

Relevance to current work: The results of this study indicated that dynamic assessment yields good classification accuracy using a combination of DA story scores, modifiability rating, and ungrammaticality from a language sample for children who are English language learners.


Objective: The purpose of this study was to investigate the reliability and classification accuracy of a dynamic assessment of narrative language. There were two experiments in the study. The first experiment examined the internal consistency of the two stimulus stories used in the dynamic assessment. The second experiment examined the classification accuracy of the dynamic assessment.

Method: The first experiment included 58 first and second-grade students who are African American, European American, and Latino American. Each student was asked to narrate two stories using wordless picture books. The stories’ story components, story ideas and language, episode structure, and productivity were measured. The second experiment included 71 first and second-grade students who are African American, European American, and Latino American. The students were placed in three groups: 27 students were in the typical development (TD) group, 14 students were in the language impairment (LI) group, and 30 students were in the no-treatment control group. Each
student was administered the dynamic assessment of narrative language using a test-teach-retest model. During the assessment, each child was asked to narrate a story using a wordless picture book as pretest. After the pretest, the examiner provided explicit instruction focusing on story components and episode structure. After the teaching phase, each child was administered a posttest using a different story.

Results: For the first experiment, the results indicated that the two stimulus stories used in the dynamic assessment had good internal consistency. For the second experiment, the results indicated that students in the TD group demonstrated greater gains from pretest to posttest than students in the LI group and no-treatment control group. The combination of modifiability and posttest scores yielded perfect classification accuracy (sensitivity and specificity) in identifying students with and without language disorders.

Relevance to current work: Dynamic assessment yields high classification accuracy in identifying students with language disorders, especially students who are from diverse cultures and language backgrounds. A combination of modifiability rating scores and posttest scores yielded the highest classification accuracy when used predictors of language ability.


Objective: The purpose of this study was to demonstrate the application of the mediated learning experience to dynamic language assessment with children who were from diverse cultural and language backgrounds.
Method: This study included 60 Spanish and English-speaking Puerto Rican and African American children between the age of 3;07 and 4;09. Some children were placed into the possibly language disordered (PLD) group based on classroom language use, clinical judgement, and parent and teacher reports. Each child was administered the Expressive One-Word Picture Vocabulary Test (EOWPVT) as pretest. After the administration of the pretest, two 20-minute intervention (teaching phase) in small groups were conducted. During the intervention, the examiners recorded the results of teaching phase by using a modified version of the Dynamic Assessment Recording Form and a three Likert-type scales modifiability rating form. The EOWPVT was administered again after the teaching phase as posttest to each child.

Results: The results indicated that there was no significantly difference between the PLD group and nondisabled group for pretest. However, there was a significantly different pattern of results for the 2 groups (PLD group and nondisabled group) from pretest and posttest. Children in the nondisabled group gained more scores than the children in the PLD group from pretest to postest. Children in the PLD group also demonstrated significantly lower modifiability rating than children in the nondisabled group.

Relevance to current work: Dynamic assessment is a more valid and accurate measure for children from diverse cultural and language backgrounds as the results from the static measure used in the study showed no significant difference between the PLD group and nondisabled group for the pretest.

https://doi.org/10.1044/2016_jslhr-l-15-0426

**Objective:** The purpose of this study was to investigate the classification accuracy of an English dynamic assessment of narratives in identifying language disorders.

**Method:** This study included 42 Spanish-English bilingual students from kindergarten to third grade. One third of the students had language disorders. Each child was administered two 25 minutes dynamic assessment of narrative language using the test-teach-test model. During the pretest and posttest, each child was asked to retell a narrative. The examiner scored both tests in real time. Between the pretest and posttest, the examiner provided structured intervention to each child targeting missing story grammar elements and language complexity. The administration of the pretest and posttest was identical, but each present a different story with similar structure and difficulty.

**Results:** Of the four classification predictors (posttest scores, gain scores, modifiability ratings, and teaching duration), discriminant function analysis indicated that the overall modifiability ratings were most predictive of identifying language disorders. There was 100% sensitivity and 88% specificity after the first dynamic assessment session and 100% sensitivity and 100% specificity after the second dynamic assessment session when using modifiability ratings as predictors of language disorders. When using any two combination of posttest scores, modifiability ratings, and teaching duration after one session, both sensitivity and specificity were over 90%. Comparing to two 5-10 minute teaching cycles, a post hoc analysis revealed that a similar classification accuracy
can be obtained with one single teaching cycle. This finding can potentially further the abbreviation of the dynamic assessment.

*Relevance to current work:* The current study is based on the dynamic assessment model used in this study. This study indicated that dynamic assessment an accurate form of language assessment with high sensitivity and specificity. Modifiability scores were most predictive of identifying language disorders. Gain scores may not be as predictive as other classification predictors (posttest scores, modifiability ratings, and teaching duration) in identifying language disorders. Furthermore, this study indicated that one single teaching phase can yield a similar classification accuracy as two teaching phases which can shorten the administer time of a dynamic assessment.


https://doi.org/10.1177/0022219413486930

*Objective:* The purpose of this study was to evaluate the predictive validity of a dynamic assessment of reading for bilingual Spanish-English speaking children at risk for language disorders.

*Method:* This study included 63 kindergarten bilingual Spanish-English speaking children with a mean age of 65.3 months. The children were placed in two groups: typical language (TL) and language difficulty (LD) based on the information from the parent questionnaires and results from the Bilingual English-Spanish Assessment (BESA). Each child was administered a dynamic assessment of non-sense word decoding. The dynamic assessment used a test-teach-test model with modifiability rating. Dynamic assessment phoneme gain score, dynamic assessment residuum gain score, decoding strategy score,
response to instruction score, and dynamic assessment modifiability score were analyzed. At the end of 1st grade, the results of criterion measures on non-sense word fluency, reading fluency, and word identification were collected from the children’s school district.

Results: The results indicated that the dynamic assessment modifiability score and the dynamic assessment residuum gain score were significantly correlated with the 1st grade criterion measures. The dynamic assessment also yielded high classification accuracy, with sensitivity and specificity at or above 80% for all the criterion measures, including 100% sensitivity for 2/3 first-grade measures.

Relevance to current work: Dynamic assessment yields high sensitivity and specificity. It has excellent classification accuracy in predicting future language difficulty while static measures have a history of not yielding adequate classification accuracy, especially for children from diverse cultural and language backgrounds.


Objective: The purpose of this study was to investigate the developmental language disorder (DLD) phonotype of Mandarin-speaking children using a narrative assessment and to identify the elicitation method of narrative production that can differentiate children who are at risk (AR) for DLD from children who have typically developed (TD) language.
**Method:** The study included 42 children. The children were placed in two groups: 21 children were in the AR group and the other 21 children were in the TD control group based on the results from a screening battery. Each child was administered the Multilingual Assessment Instrument of Narrative (MAIN). During the assessment, each child was asked to produce a story-retell followed by a story-tell.

**Results:** The results indicated that children in the AR group demonstrated difficulty in story macrostructure, lexical diversity, and sentence complexity. However, grammaticality and productivity were relatively reserved. Children in the AR group benefited from the adult model in the story-retell task by producing comparable numbers of story grammar elements and multiclausal utterances in the story retells as the children in the TD group. However, their story grammar elements and utterances complexity significantly decreased as the adult model was withdrawn while children from the TD group maintained their performance across tasks.

**Relevance to current work:** The phonotypes of Mandarin-speaking children who are at risk or have DLD are difficulty in story macrostructure, lexical diversity, and sentence complexity. The use of story retells revealed that Mandarin-speaking children were less likely to show and maintain gains from adult models and teaching which is one of the elements of dynamic assessment.


**Objective:** The purpose of this study was to identify the narrative abilities of Mandarin-speaking children with and without language disorders and investigate whether narrative
measures can differentiate children have language disorders from children who have typically developed language.

*Method:* The study included 36 Mandarin-speaking children between the age of 4;11 and 5;10. The children were placed in two groups: 18 were in the language disorder group and the other 18 were in the typically developed language control group based on the results of a norm-referenced language assessment. Each child was asked to produce a narrative sample using a wordless picture book “Frog, where are you?”. The narrative samples were analyzed at the macrostructure and microstructure levels.

*Results:* The results indicated that children in the language disorder group produced significantly fewer story grammar elements, evaluative comments, and coherence than children in the control group. Children in the language disorder group also demonstrated significantly less variety of words in their narrative samples. However, there was no significant difference in story length, syntactic complexity, and use of conjunctions between the two groups.

*Relevance to current work:* This study revealed that Mandarin-speaking children with language disorders demonstrated difficulty at the macrostructure levels and lexical diversity in narratives. And the utility of macrostructure and lexical diversity in a narrative assessment can identify children with language disorders.

Objective: The purpose of this study was to examine the use of dynamic assessment to evaluate the language-learning ability of Native American kindergartners.

Method: 23 Native American kindergarteners participated in this study. Based on teacher reports and classroom observation, 15 participants were identified as stronger language learners and 8 participants were identified as weaker language learners. Via a test-teach-test protocol, the participants were tested in categorizing within a 3-week period. During the teaching phase, the participants’ responsiveness to learning was measured using a modifiability index. The modifiability index was a combined score to reflect each participant’s learning strategies such as ability to attend, plan, self-regulate, and respond to learning situations.

Results: The results showed that the participants who were identified as stronger language learners demonstrated significantly greater modifiability and post-test scores than the participants who were identified as weaker language learners. All participants demonstrated gains after the teaching phase. However, participants who were identified as stronger language learners made more gains than the participants who were identified as weaker language learners.

Relevance to current work: This study indicated that dynamic assessment was a reliable tool to access the language learning ability of Native American children.

**Objective:** The purpose of this study was to examine the psychometric properties of a newly developed receptive and expressive language assessment in Mandarin called the Mandarin Clinical Evaluation of Language for Preschooler’s Core Scale (MCELP-CS).

**Method:** 379 preschool-aged participants, including 81 children with language disorders participated in this study. All participants and their caregivers were native Mandarin speakers. The participants were assessed using the newly developed MCELP-CS which consisted of 5 subscales: vocabulary comprehension, sentence comprehension, vocabulary naming, sentence structure imitation (SSI), and story narration. The participants were also assessed using the Peabody Picture Vocabulary Test-revised (PPVT-R) to examine the convergent validity of the MCELP-CS.

**Results:** The results of this study indicated that the MCELP-CS demonstrated high internal consistency and convergent validity. It also indicated good sensitivity (.89) and specificity (.89).

**Relevance to current work:** This study provided information on the different kinds of language assessments that were commonly used for screening and diagnosing children with language disorders in Mandarin. According to this study, most of the language assessments used in China were norm-referenced standardized tests that only include children who were Mandarin native speakers from the mainstream Han culture. Even though the assessment used in this current study appears to have good psychometric properties with children who were Mandarin native speakers from the mainstream Han culture, there could be bias towards children who are Mandarin language learners and children from another culture. There is a need to reduce cultural and linguistic bias in Mandarin language assessments.
APPENDIX B

Institutional Review Board (IRB) Approval Letter

Memorandum

To: Douglas Petersen
Department: BYU - EDUC - Communications Disorders
From: Sandee Aina, MPA, HRPP Associate Director
Wayne Larsen, MAcc, IRB Administrator
Date: March 31, 2022
IRB#: IRB2020-328
Title: Examining the Validity and Reliability of Dynamic Assessments of Reading and Language

Brigham Young University’s IRB has approved with conditions, the research study referenced in the subject heading as exempt level, categories 1 and 2. The study cannot occur until you have received specific research site approvals from the school districts in Utah and the approval of schools and districts outside of Utah.

This study does not require an annual continuing review. Each year near the anniversary of the approval date, you will receive an email reminding you of your obligations as a researcher and to check on the status of the study. You will receive this email each year until you close the study.

The study is approved as of 03/31/2022. Please reference your assigned IRB identification number in any correspondence with the IRB.

Continued approval is conditional upon your compliance with the following requirements:

1. A copy of the approved informed consent statement can be found in iRIS. No other consent statement should be used. Each research subject must be provided with a copy or a way to access the consent statement.
2. Any modifications to the approved protocol must be submitted, reviewed, and approved by the IRB before modifications are incorporated in the study.
3. All recruiting tools must be submitted and approved by the IRB prior to use.
4. Instructions to access approved documents, submit modifications, report adverse events, can be found on the IRB website, iRIS guide: https://irb.byu.edu/iris-training-resources
5. All non-serious unanticipated problems should be reported to the IRB within 2 weeks of the first awareness of the problem by the PI. Prompt reporting is important, as unanticipated problems often require some modification of study procedures, protocols, and/or informed consent processes. Such modifications require the review and approval of the IRB. Please refer to the IRB website for more information.
### APPENDIX C

**Study 1 Story**

<table>
<thead>
<tr>
<th>Character/Setting</th>
<th>Problem</th>
<th>Feeling</th>
<th>Plan/Attempt</th>
<th>Consequence/End emotion</th>
</tr>
</thead>
</table>
| 有一天，东尼在公园里骑自行车。 | 他意外地撞到石头。他摔倒了。 | 他很伤心因为他受伤了。 | 于是，他决定找妈妈来帮助他。 | 妈妈帮助他和给他一个绑带。之后，他感觉好多了。
他很开心因为他可以骑单车了。 |

“One day Tony was riding his bike at the park.”

“He accidentally crashed into a rock and fell.”

“He was sad because he was hurt.”

“Then, he decided to get help from his mom.”

“His mom helped him and gave him a band-aid. After that, he felt much better. He was happy because he could go back and ride his bike.”

Adapted from PEARL (Petersen & Spencer, 2014)
## Study 1 Scoring

<table>
<thead>
<tr>
<th>Story Grammar (SG)</th>
<th>2 points</th>
<th>1 point</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Character</strong></td>
<td>东尼/任何名字 (Tony or any other names)</td>
<td>男孩 (Boy)</td>
</tr>
<tr>
<td><strong>Setting</strong></td>
<td>在公园骑自行车 (Biking at the park)</td>
<td>骑自行车/公园 (Biking/park)</td>
</tr>
<tr>
<td><strong>Problem</strong></td>
<td>摔倒/跌倒 (Fell down)</td>
<td>受伤 (Got hurt)</td>
</tr>
<tr>
<td><strong>Feeling</strong></td>
<td>伤心/生气 (Sad/mad)</td>
<td>不喜欢它/哭 (Didn’t like it/cry)</td>
</tr>
<tr>
<td><strong>Plan</strong></td>
<td></td>
<td>决定/想 (Decided/thought)</td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td>找妈妈来帮助他 (Found mom to help him)</td>
<td>告诉妈妈 (Told mom)</td>
</tr>
<tr>
<td><strong>Consequence</strong></td>
<td>妈妈给他绷带/邦迪 (Mom gave him a bandage)</td>
<td>妈妈帮助他 (Mom helped him)</td>
</tr>
<tr>
<td><strong>Ending</strong></td>
<td>骑自行车 (Biking again)</td>
<td>好多了/没事了 (He felt better)</td>
</tr>
<tr>
<td><strong>End Feeling</strong></td>
<td>开心 (Happy)</td>
<td>感觉好多了 (Felt much better)</td>
</tr>
</tbody>
</table>

**Story Grammar Subtotal (SG):** 16

### Language Complexity (LC)

<table>
<thead>
<tr>
<th>Word</th>
<th># of times used</th>
</tr>
</thead>
<tbody>
<tr>
<td>之后 (After)</td>
<td>1</td>
</tr>
<tr>
<td>因为 (Because)</td>
<td>2</td>
</tr>
<tr>
<td>于是 (Then)</td>
<td>1</td>
</tr>
</tbody>
</table>

**Language Complexity Subtotal (LC):** 4

**Episode (E)**
<table>
<thead>
<tr>
<th>Combination</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>P+A or P+C or A+C</td>
<td>2</td>
</tr>
<tr>
<td>P+C+E</td>
<td>3</td>
</tr>
<tr>
<td>P+A+E</td>
<td>3</td>
</tr>
<tr>
<td>P+A+C</td>
<td>4</td>
</tr>
<tr>
<td>P+A+C+E</td>
<td>5</td>
</tr>
</tbody>
</table>

**Episode Subtotal (E):** /5

**Pretest/Posttest Total Score (SG+LC+E):** /25

Adapted from PEARL (Petersen & Spencer, 2014)
APPENDIX E

Study 1 Pictures of Story Used in Teaching Phase

PEARL (Petersen & Spencer, 2014)
APPENDIX F

Study 1 Story Grammar Icons Used in Teaching Phase

PEARL (Petersen & Spencer, 2014)
APPENDIX G

Study 1 Instruction and Script for Teaching Phase

步骤1 (Step 1):
称赞 (Praise) 说 - “现在，我会帮您把故事讲得更好。” (Say- “Now, I am going to help you tell even better stories.”)
显示图片。阅读故事，并指向相应的图标。 (Show Pictures. Read the story and point to the corresponding icons.)
阅读故事后，指向并命名图标并提供故事的示例。 (指向角色图标，并说： “这图标代表‘角色’。这个故事中的角色是东尼。”) (After reading the story, point to and name the icons and provide example from the story. (Point to the character icon and say, “This means ‘character’. The story in this character in this story is Tony.”)

图标：角色、场景、问题、感觉、动作、结尾、结束时的感觉
(Icons: Characters, Setting, Problem, Feeling, Action, Ending, End feeling)

步骤2 (Step 2):
说 - “这次你要把这个故事讲给我听。我会帮你。” (Say- “This time you are going to retell the story. I will help you with it.”)
显示图片。必要时，使用1级和/或2级提示来帮助学生重述故事的所有部分。 (Show Pictures. Help students retell all parts of the story using level 1 and/or level 2 prompts when necessary.)

步骤3 (Step 3):
称赞 (Praise) 说 - “您要把这个故事再讲给我听。但是这次没有图片。如果您需要的话，我会帮你。” (Say “Now I am going to take away the pictures and you are going to retell this story. I will help you if you need it.”)
仅显示图标。必要时使用1级和/或2级提示帮助学生复述故事的所有部分。 (Show only the icons. Help students retell all parts of the story using level 1 and/or level 2 prompts when necessary.)

步骤4 (Step 4):
称赞 (Praise) 说 - “您要把这个故事再讲给我听，但是这次没有图片也没有图标。如果您需要的话，我会帮你。” (Say “This time, you are going to tell the story without the icons. I’ll help you if you need)
在必要时，使用1级和/或2级提示来帮助学生重述故事的所有部分。 (Show Pictures. Help students retell all parts of the story using level 1 and/or level 2 prompts when necessary.)

1级提示 (Level 1 Prompt)：开放式问题 (例如“他的问题是什么？” / “他做了什么来解决问题？” Open ended questions (e.g., “What was his problem?” / “What did he do to fix his problem?”)
2 级提示(Level 2 Prompt): 示范学生应该说什么，例如“他很伤心，因为他的手肘在流血。”Model what the student should say (e.g. “He was sad because his elbow was bleeding.”)

Adapted from PEARL (Petersen & Spencer, 2014)
### Study 1 Modifiability Rating

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>反应</strong> Responsiveness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>提示Prompts</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>信心Confidence</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>中断Disruptions</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>速度Rate</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>学习Learning</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

总分数(Total Responsiveness Score)： /4

Adapted from PEARL (Petersen & Spencer, 2014)
APPENDIX I

Study 1 Teacher Rating

教师评分 (Teacher Rating)
在 1-10 的范围内，学生目前理解中文的能力是多少？ __________
On a scale of 1-10, what is the student’s current ability to understand/comprehend Chinese?
### Study 2 Story

<table>
<thead>
<tr>
<th>Character/Setting</th>
<th>Problem</th>
<th>Feeling</th>
<th>Consequence/End emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td>昨天，小伟在公园骑车。</td>
<td>他摔倒了。</td>
<td>他难过。</td>
<td>妈妈帮小伟。之后，他很开心因为他可以骑车了。</td>
</tr>
<tr>
<td>&quot;Yesterday, Xiaowei was riding his bike at the park.&quot;</td>
<td>&quot;He fell.&quot;</td>
<td>&quot;He was sad.&quot;</td>
<td>&quot;Mommy helped Xiaowei. After that, he was happy because he could ride his bike.&quot;</td>
</tr>
</tbody>
</table>

Adapted from PEARL (Petersen & Spencer, 2014)