New Approaches to C-test Development:
Examples from Mandarin, Turkish, and Russian

Margaret E. Malone, Georgetown University, ACTFL
Yiran Xu, Seyma Toker, Derek Reagan, Georgetown University
AGENDA

• C-test Background
• New Features in the AELRC C-test Projects
• Current AELRC C-test Projects
  • Mandarin
  • Turkish
  • Russian
• Conclusions & Future Directions
C-TESTS

• Similar to a cloze test in form
• Deletion method: 2-2-2 rule
  • Delete the second half of every second word, beginning from the second sentence (Raatz & Klein-Braley, 1981)
• 25 blanks for each text, 5 texts in total
• Quick and cost-effective
  • Open access; around 30mins
  • Placement tests (Mozgalina & Ryshina Pankova, 2015; Norris, 2006; Wilmes, 2007)
• Good indication of global or integrated language proficiency (e.g., Eckes & Grotjahn, 2006; Babaii & Ansari, 2001; Babaii & Jalali Moghaddam, 2006)
PREVIOUS C-TEST RESEARCH

Letter-based scripts

• English (e.g., Baghaei, 2008; Babaii & Ansary, 2001; Harsch & Hartig, 2015; Jafarpur, 1995)
• French (Grotjahn & Stemmer, 1985; Reichert, Keller & Martin, 2010; Ward, 1987)
• German (Eckes & Grotjahn, 2006; Mozgalina & Ryshina-Pankova, 2015; Norris, 2006; Wilmes, 2007)
• Portuguese (Maimone, 2018)
• Turkish (Demiralp, 2018)

Character-based scripts

• Korean (Lee-Ellis, 2009; Son, Kim, Cho & Davis, 2018)
• Japanese (Sasayama, 2018)
• Bangla (McKay & Abedin, 2018)
• Mandarin (Arras & Grotjahn, 1994; Zhang, 1985)
C-TEST DEVELOPMENT & VALIDATION PROCEDURE

1. TEXT SELECTION
2. EXPERT REVIEW
3. TEST CREATION
4. NATIVE SPEAKER PILOT
5. TEXT REVISION
6. PILOT ON LEARNERS
7. TEST VALIDATION
NEW FEATURES IN OUR C-TEST PROJECTS

Think-aloud (Mandarin, Turkish, & Russian)

Two Deletion Methods: Word-based & Stoke-based (Mandarin)

听 → 聽

Two versions available: Simplified & Traditional (Mandarin)

Text selection based on ACTFL Guidelines (Turkish)

Validation with ACTFL tests (Mandarin)
THE C-TEST PROJECTS AT AELRC

• Japanese (Sasayama, 2018)
• Korean (Son, Kim, Cho, & Davis, 2018)
• Bangla (McKay & Abedin, 2018)
• Mandarin (Xu & Malone, in preparation)
• Turkish, English, and Russian (ongoing projects)

Doctoral Dissertations
• Korean C-test for heritage language learners (Son, 2018)
• A meta-analysis of C-test studies (McKay, 2019)
The Mandarin C-test Project
No. 1.

我的名字叫马克，我来自美国纽约，我来中国已经三年了。我现(1)是北京大(2)三年(3)的学生，我在(4)学习中(5)和国(6)关系。我喜(7)旅行，美(8)，运动和电(9)。周末的时(10)我喜欢跟我的中国朋(11)一起打打篮(12)，看看电影，或(13)在北京附(14)玩一玩，品(15)地方的小(16)。我也很喜欢跟朋友去酒(17)聊天喝(18)，认识更多新的朋友。我很想(19)我的家(20)，我每(21)都会跟他们打电(22)。我的爸爸妈妈告(23)他们想来中国看我，我很高(24)。到时候，我会带他们去爬长城，看故宫，吃一些北京的特(25)小吃。我已经等不及要见到他们了！
MANDARIN LANGUAGE

• Non-alphabetic; based on a logographic system
• Constructed horizontally from left to right
• No spacing between words and characters

• What is a word in Chinese?
  • A minimal linguistic form that has meaning, can occur independently in speaking and writing (Fu, 1985; Zhu, 1982) and has translation equivalent in other languages (Chao, 1968) \( \rightarrow \) (therefore, particles like “ba” “le” “ma” do not count as words)
  • It may include 1 character 我 (I); or 2 characters 旅行 (travel) or more
TEXT SELECTION (20 texts)

- A variety of familiar genres

- All from authentic sources (included texts from textbook for school-age children for lower-levels)

- Generic topics

- Self-contained

- Texts were selected based on the descriptors of HSK (Hanyu Shuiping Koashi; translated as the Chinese Proficiency Test) administered by Hanban

NEWSPAPER

- Text 1, 6, 8, 19
- Entertainment, life style, social network, culture

MAGAZINE

- Text 13, 14, 17, 18
- Health, anthropology, travel, communication

POPULAR MEDIA

- Text 2, 3, 5, 7
- Commentary, travel descriptions, product description, customer review

BLOGS, LETTERS, EMAILS

- Text 4, 15, 16, 20
- Education, politics, family, cover letter

TEXTBOOK FOR SCHOOL-AGE CHILDREN

- Text 9, 10, 11, 12
- Attitude, philosophy, animal, travel
EXPERT REVIEW

• 25 Chinese instructors with 3+ years teaching experience (average: more than 10 years)

• Teaching level: university (21), secondary (1), elementary (1), mixed (2)

• Recruited from more than 100 U.S. universities

• Education: Master’s (12), PhD (13)

Survey

• Rate the difficulty level of 20 texts (1-5, 1 being very easy, 5 being very difficult)

• Provide qualitative feedback

• Eliminated 5 texts (out of 20)
  • Technical
  • Culture-specific
  • Required background knowledge

• Added two lower-level texts (manipulated texts)
  • Self-introduction
  • Narrative of daily activities
PHASE 2 TEST CREATION & NATIVE SPEAKER PILOT
TWO DELETION METHODS

Word-based deletion: delete the second half of every other word from the second sentence

Stroke-based deletion: delete the second half (by the order of the stroke) of every other character from the second sentence

(15) 我的名字叫马克，我来自美国纽约，我来中国已经三年了。我现在是北京大学三年级的学生，我在这学习中文和国关系。我喜欢旅行，美食，运动和电影。周末的时候我喜欢跟我
Deletion Rules & Scoring Principles

- Manual checking

What is not deleted:
- Numbers
- Single-stroke characters
- Single-character words
- Repetitive characters and words
- Proper nouns

Accept alternative answers
Accept both simplified and traditional characters
Accept orthographic errors if they do not interfere with meaning

Scored dichotomously (0 or 1)
Score range for each text: 0 - 25
NATIVE SPEAKER PILOT

GROUP 1: Word-based
- 30 Mandarin native speakers living in the U.S. and China
- Age: $M = 21.6; SD = 2.89$
- Think-aloud (N=4)
- Undergraduate and graduate students
- Provided feedback on text difficulty

GROUP 2: Stroke-based
- 31 Mandarin native speakers living in the U.S. and China
- Age: $M = 25.56; SD = 3.34$
- Think-aloud (N=4)
- Undergraduate and graduate students
- Provided feedback on text difficulty
NS PILOT RESULTS

- Range: 23.6 – 25
- Average accuracy rate:
  - Word-based: 98.87%
  - Stroke-based: 97.92%
- More variations in the stroke-based version
- Think-aloud & qualitative feedback:
  - The stroke based is more challenging because it does not allow for alternative answers
  - Sometimes, no contextual information is needed to complete the character for the stroke-based version
PHASE 3 TEXT REVISION & LEARNER PILOT
TEXT REVISION

- Word-based deletion
- Deleted another 5 texts (out of 15)
  - Challenging for NSs
  - Translated articles
- Revised the remaining texts (based on feedback & think-aloud data)

TEXT TOPICS

- TEXT 1: Manipulated text: Self-introduction
- TEXT 2: Manipulated text: Visiting a friend in hospital
- TEXT 4: Cover letter: Job application
- TEXT 5: Customer Reviews: Bubble tea
- TEXT 6: Opinion: Livestreaming industry in China
- TEXT 7: Opinion: The philosophy of growing flowers
- TEXT 8: Description: Traveling in Tibet
- TEXT 9: Description: Avocado
- TEXT 10: Comments: Museums in China
LEARNER PARTICIPANTS

• Computer-based (or Paper-based test upon request) (40-50mins) + background questionnaire (5mins)

Learners of Chinese (N = 34)
• Age: Mean 22.9 (SD 5.11)
• Gender: 12 M; 22 F
• Recruited from 3 major universities in the U.S.
• 19/34 (Currently enrolled in a Chinese program)
• Think-aloud (N = 21)
• ACTFL OPIc & RPT (N = 24)

Self-reported proficiency (on a scale of 0 to 100)
• Reading (M = 41.74, SD = 29.78, Range 0-91)
• Writing (M = 38.35, SD = 28.91, Range 0-80)
• Listening (M = 43.09, SD = 29.07, Range 1-100)
• Speaking (M = 35.59, SD = 26.97, Range 1-80)

Average self assessment in four levels
• Level 1: 10 (0-25)
• Level 2: 12 (25-50)
• Level 3: 7 (50-75)
• Level 4: 5 (75-100)
DESCRIPTIVE STATISTICS

- Possible range: 0-25
- Actual range: 0-25
- Mean scores: 5.8 – 17.35
- Cronbach’s alpha: 0.98
Summary fit statistics within range

- Person separation index: 7.73
- Separation reliability: 0.98
- Real RMSE: 0.18

- Item separation index: 7.73
- Separation reliability: 0.99
- Real RMSE: 0.09

Super-items: scale 0-25
Rating Scale Model
SELECTED FIVE TEXTS RESULTS

Summary fit statistics within range

- Person separation index: 6.52
- Separation reliability: 0.94
- Real RMSE: 0.32
- Item separation index: 13.14
- Separation reliability: 0.99
- Real RMSE: 0.10
CORRELATIONS: C-TEST & ACTFL RPT RATINGS

Reading Proficiency Test & C-test logit scores on 5 selected texts:
\[ r = 0.91 \ (p < 0.01, \ n = 24) \]
CORRELATIONS: C-TEST & ACTFL OPIc RATINGS

OPIc & C-test logit scores on 5 selected texts:
\( r = 0.81 \) (\( p < 0.01, \) n = 24)
CORRELATIONS: C-TEST & SELF-ASSESSMENT

Reading ($r = 0.85, p < 0.01$)  
Listening ($r = 0.85, p < 0.01$)  
Speaking ($r = 0.75, p < 0.01$)  
Writing ($r = 0.75, p < 0.01$)
DISCUSSION

• A reliable tool for SLA research purposes
  • Quick (30 mins) and free (upon request)
  • High reliability indices
  • Distribute learners into different major levels
  • Measure global language proficiency
    • Highly correlated with ACTFL RPT and OPIc ratings
    • Highly correlated with self-assessed proficiencies in four skills

• However...
  • Since Chinese is a logographic-based language, it is challenging for learners with limited literacy skills to meet basic demands of the test
  • Correlate more with receptive than productive language skills
The Turkish C-test Project
THE TURKISH C-TEST PROJECT

Goal
• Revise the previous Turkish C-test (Demiralp, 2018), which was designed according to ILR reading scale, to align with the ACTFL proficiency guidelines

New Features
• ACTFL guidelines for text selection
• Holistic rubric for expert review
• Think-aloud protocol in interviews with experts
PHASE 1 TEXT SELECTION & EXPERT REVIEW

TEXT SELECTION
• 11 validated texts from Demiralp (2018)

EXPERT REVIEW
• Survey and interview with 10-15 Turkish language instructors

Text 1

14. In your opinion, what is the proficiency level of this text?
   _Novice
   _Intermediate
   _Advanced
15. Now, read the items under each level and check all that apply.

<table>
<thead>
<tr>
<th>Novice</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>The text includes key words, cognates, and formulaic phrases that are highly contextualized.</em></td>
<td><em>The text is primarily organized in individual sentences and strings of sentences containing predominantly high frequency vocabulary.</em></td>
<td><em>The text provides a straight-forward sequencing of real-world events.</em></td>
</tr>
<tr>
<td><em>The information in the text is highly predictable.</em></td>
<td><em>The information in the text conveyed in simple, predictable, loosely connected manner.</em></td>
<td><em>The information in the text is conveyed through main idea and supporting details, and sentences are neatly connected through conjunctions and transitional expressions.</em></td>
</tr>
<tr>
<td><em>The text is about a highly familiar topic that is relevant to daily life.</em></td>
<td><em>The text is about a highly familiar topic that is relevant to daily life.</em></td>
<td><em>The text is uncomplicated and the subject matter pertains to real-world topics.</em></td>
</tr>
<tr>
<td><em>The readers can easily draw on their background knowledge to understand the text.</em></td>
<td><em>The text is not complex and has a predictable pattern of presentation.</em></td>
<td><em>The comprehension of the text is supported by the reader’s knowledge of the conventions of the language and Turkish prose.</em></td>
</tr>
<tr>
<td><em>The text includes words with letters unique to Turkish alphabet (e.g. “ç, ş, ğ, ö, ü”)</em></td>
<td><em>The texts is primarily written in present and present progressive tense.</em></td>
<td><em>The text includes more complex grammatical structures such as passive voice, past and future tense.</em></td>
</tr>
</tbody>
</table>
NEXT STEPS IN THE TURKISH C-TEST PROJECT

1. TEXT SELECTION
2. EXPERT REVIEW
3. TEST REVISION
4. NATIVE SPEAKER PILOT (20-50 native speakers)
5. TEST REVISION
6. LANGUAGE LEARNER PILOT (20 non-native speaker per level)
7. TEST VALIDATION
The Russian C-test Project
THE RUSSIAN C-TEST PROJECT

Goal
• To develop valid and reliable Russian language C-tests aligned with the American Council on Teaching of Foreign Languages (ACTFL) and Interagency Language Roundtable (ILR) proficiency guidelines

New Features
• Russian language C-test
• Subject-matter expert passage evaluations
• Think-aloud protocols for NSs and NNSs
RUSSIAN LANGUAGE FEATURES

Challenges
• Cyrillic alphabet
• Highly fusional morphology
• Verb conjugation for person, number, tense, voice, and mood
• Accent stress marker (´) can distinguish between two different words
RESEARCH STAGES

**Phase 1: Text Selection**
- Texts selected by Russian language and assessment expert

**Phase 2: Expert Review**
- 30 Russian language instructors
- Text evaluation
- Deletion strategies

**Phase 3: Test Creation**
- 15 texts
- Using expert feedback for deletion methods

**Native Speaker Pilot**
- 50 NSs
- Test Pilot
- Think-aloud protocols

**Text Revision**
- 15 revised texts
- Revising deletion methods

**Pilot Pilot**
- 100 NNSs
- General language proficiency tests
- Think-aloud protocols
CONCLUSIONS & FUTURE DIRECTIONS

• Continuing research
  • Additional languages

• Explore uses (remember ILTA Code of Ethics and Guidelines for Practice)
  • Research
  • Screening

• AELRC C-test Repository (in progress)
  • Bangla, French, Japanese, Portuguese, Spanish, Arabic, Turkish
  • To be made available to foreign language researchers and instructors affiliated with a university, educational or research institution
  • Must keep C-tests both available and relatively secure
THANK YOU! 谢谢！

Teşekkür ederim! Спасибо!

Margaret Malone: malonem@georgetown.edu
Yiran Xu: yx110@georgetown.edu
Seyma Toker: st977@georgetown.edu
Derek Reagan: dr862@georgetown.edu