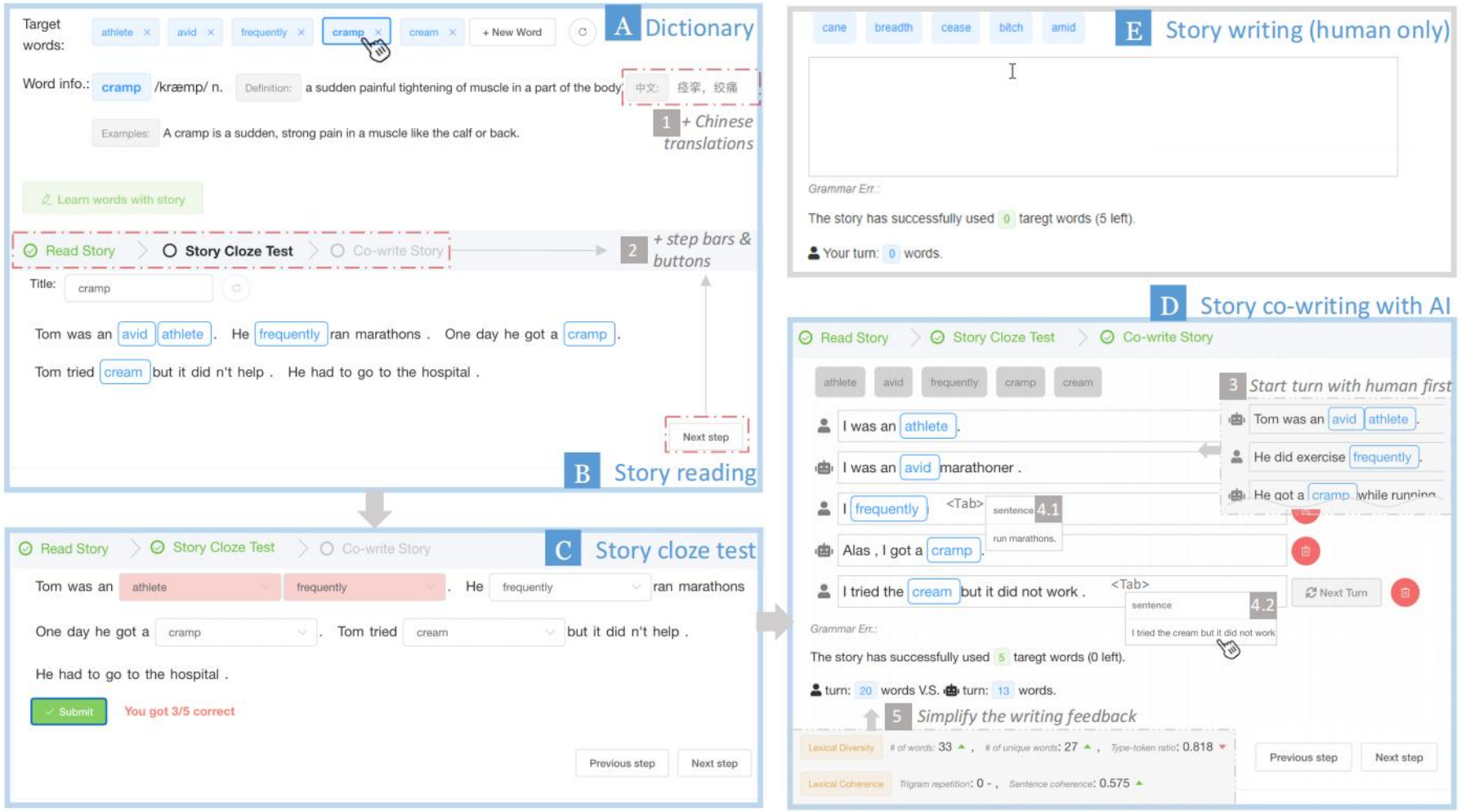
hci lab
SEOUL NATIONAL UNIVERSITY

Storyfier: Exploring Vocabulary Learning Support with T ext Generation Models

Zhenhui Peng, **Xingbo Wang**, Qiushi Han, Junkai Zhu, Xiaojuan Ma, Huamin Qu

ACM UIST

2023-11-23


A Dictionary
Target words:
athlete
avid
frequently
cramp
cream
+ New Word
Word info.:
cramp
/kræmp/ n.
Definition:
a sudden painful tightening of muscle in a part of the body
中文:
痉挛，绞痛
Examples:
A cramp is a sudden, strong pain in a muscle like the calf or back.
1 + Chinese translations
Learn words with story
Read Story
Story Cloze Test
Co-write Story
2 + step bars & buttons
Title:
cramp
Tom was an avid athlete . He frequently ran marathons . One day he got a cramp .
Tom tried cream but it did n't help . He had to go to the hospital .
Next step
B Story reading
Read Story
Story Cloze Test
Co-write Story
C Story cloze test
Tom was an athlete frequently . He frequently ran marathons
One day he got a cramp . Tom tried cream but it did n't help .
He had to go to the hospital .
Submit
You got 3/5 correct
Previous step
Next step
cane
breadth
cease
bitch
amid
E Story writing (human only)
Grammar Err.:
The story has successfully used 0 taregt words (5 left).
Your turn: 0 words.
D Story co-writing with AI
Read Story
Story Cloze Test
Co-write Story
athlete
avid
frequently
cramp
cream
3 Start turn with human first
I was an athlete .
Tom was an avid athlete .
I was an avid marathoner .
He did exercise frequently .
He got a cramp while running
I frequently <Tab>
sentence 4.1
run marathons.
Alas , I got a cramp .
I tried the cream but it did not work . <Tab>
sentence 4.2
I tried the cream but it did not work
Next Turn
Grammar Err.:
The story has successfully used 5 taregt words (0 left).
turn: 20 words V.S. turn: 13 words.
5 Simplify the writing feedback
Lexical Diversity
# of words: 33 , # of unique words: 27 , Type-token ratio: 0.818
Lexical Coherence
Trigram repetition: 0 - , Sentence coherence: 0.575
Previous step
Next step

Motivaion

- Existed tools fall short in 2 aspects:

Only leverage existing materials → lack of flexibility

Use language receptively → lack of prodctive task to master usage of words

Design & develop phases

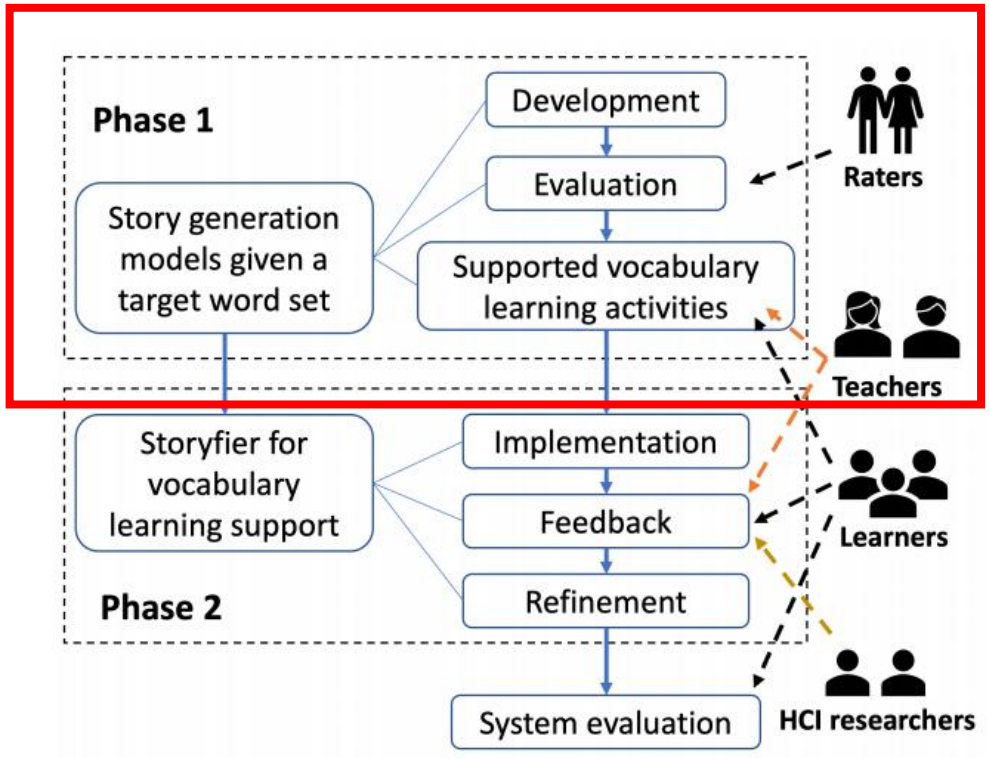


Figure 1: Our two-phases design and development process of *Storyfier* with teachers, learners, and HCI researchers.

In the whole 2-phases develop procedure, phase 1 mainly consists of 3 steps:

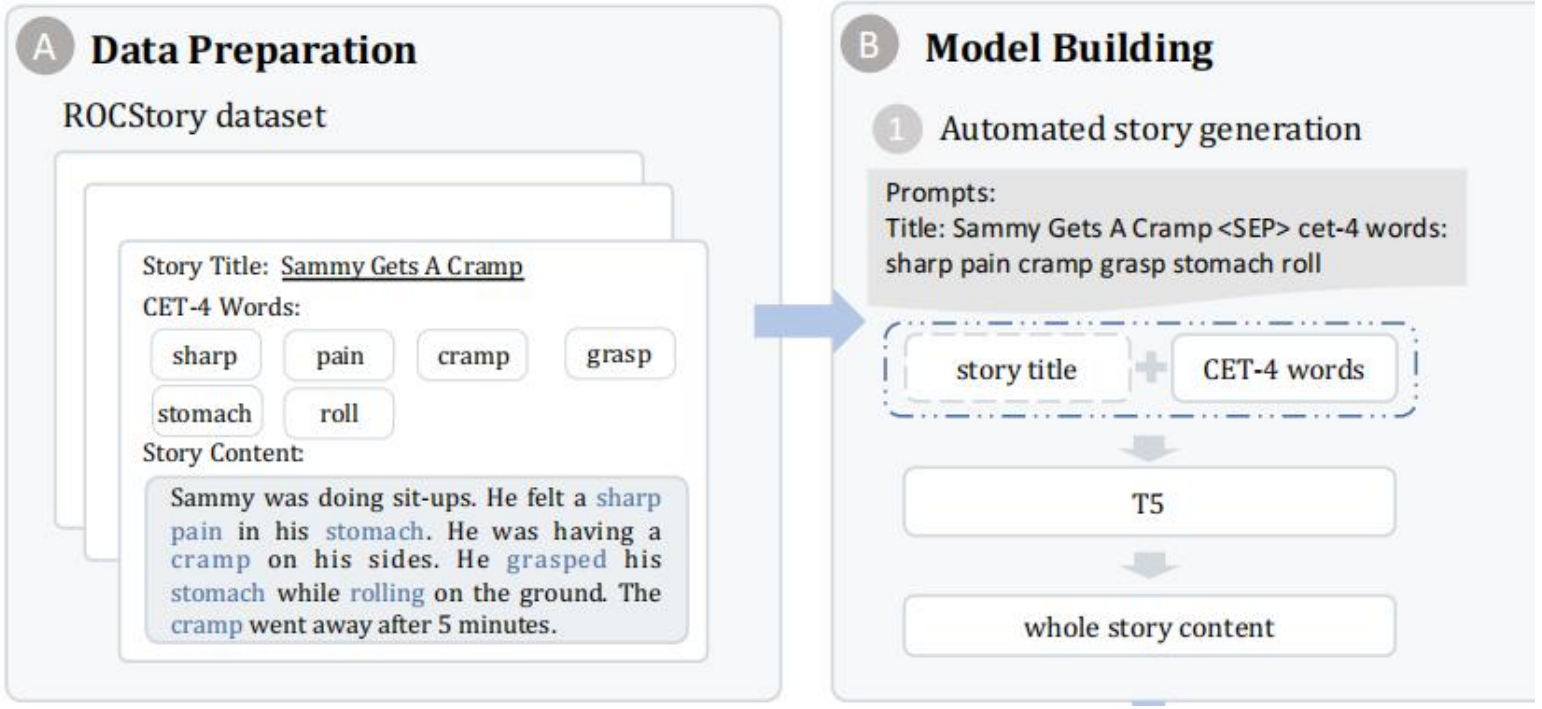
- develop **story generation model**;
- **validation** of the model;
- **explore learning activities** supported by this m odel.

Model development

- **Target: controllable model which can generate stories with target word sets**

Attributes	Values
# of stories	101,661
# of words	4,640,319
Average story length	45.65
Average sentence length	7.80
Average readability	57.14
Coverage of CET-4 words	89.52%

tuple:{ *story title*, *target words*, *story sentences* }



validate the quality of generated stories

Model validation

- **Validating object**: 20 human-machine story pairs

  20 stories from ROCstory with varied difficulty levels of contained words

  20 stories generated by model based on corresponding story title and contained words

- **Validation methods**: technical and human

Automatically evaluated →

	Grammar	Type-token ratio	Trigram repetition	Sentence coherence
Human	1.00	0.75	0.01	0.42
Machine	1.00	0.77	0.01	0.43

Evaluated by English skil led people →

	Coherence *	Relevance *	Interestingness	Overall
Human	4.53	4.58	4.08	4.26
Machine	3.92	4.33	3.97	4.01

→ **Adequate**

conduct interview to explore learning activities that can be supported by the model

Explore learning activities

- Explore possible activities by conducting **semi-structured interviews**
- **interviewees:** skilled English **teacher** and well-experienced **students** i n using vocabulary learning tools

Interact with rough interf ace to try generative mo del and brainstorm

Providing ideas of leve raging generative mod el to support vocabula ry learning

- Story reading
- Cloze test
- Turn-taking writing

Design & develop phases

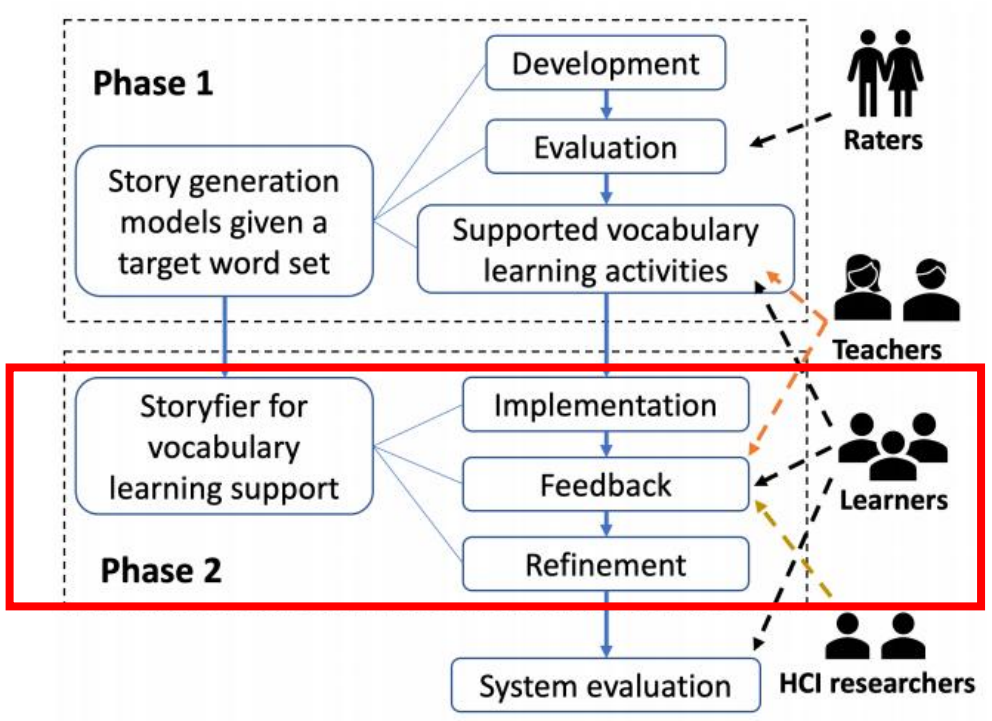
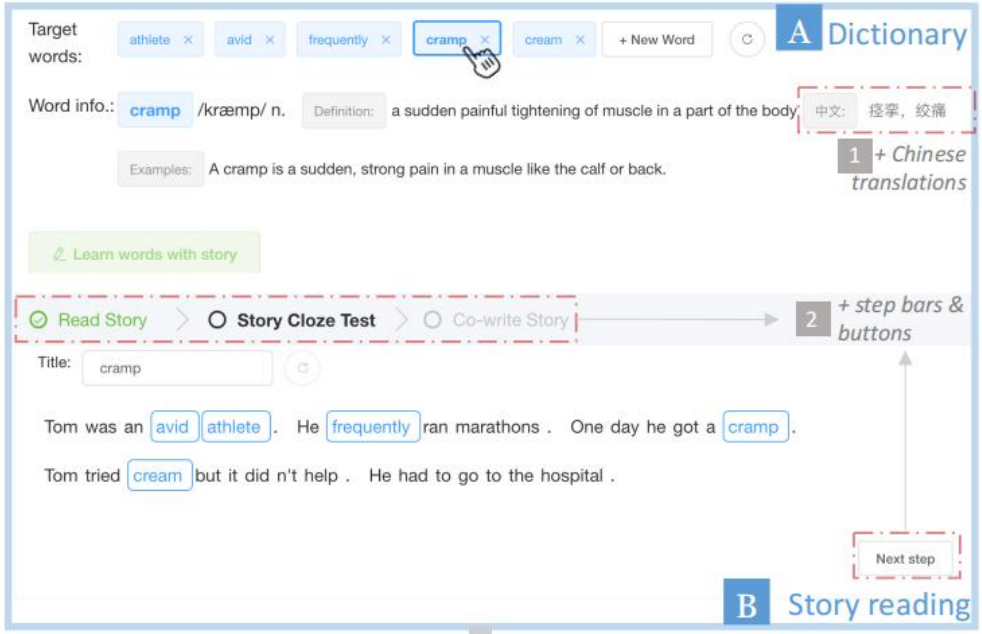


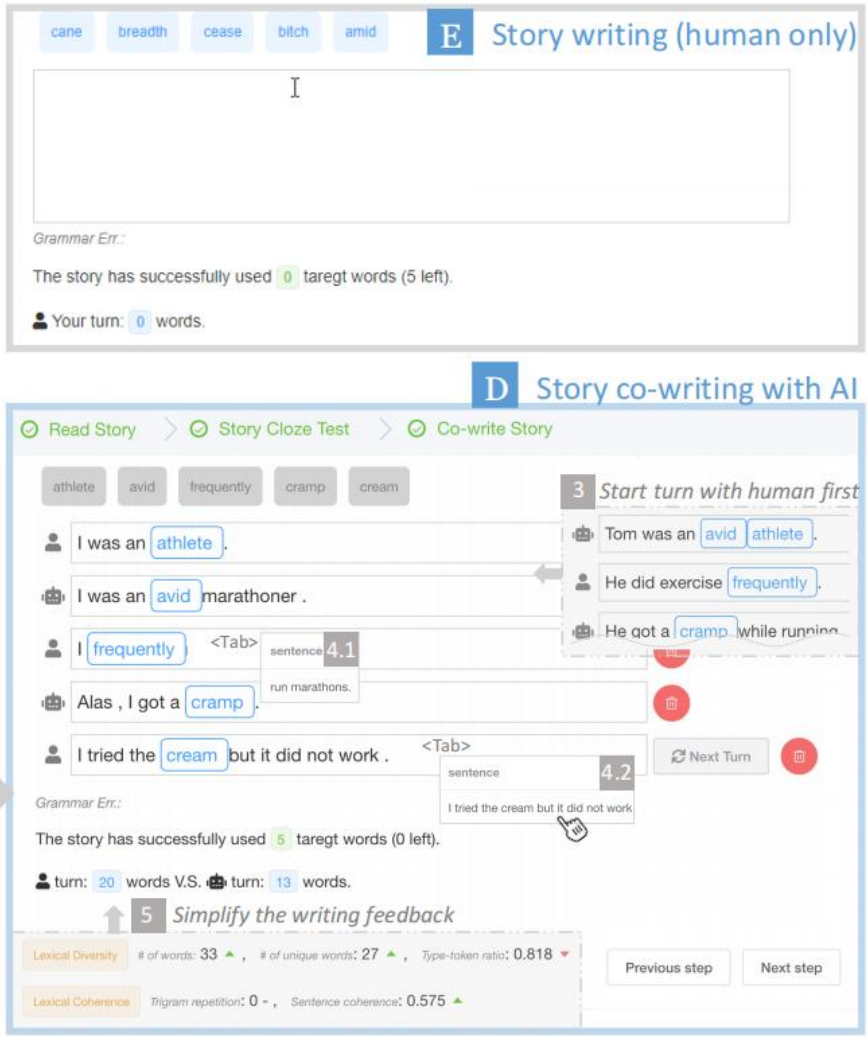
Figure 1: Our two-phases design and development process of *Storyfier* with teachers, learners, and HCI researchers.

In the whole 2-phases develop procedure, phase 2 mainly consists of 3 steps:

- system **implementation**;
- tests for **feedback**;
- **refinement** according to feedback.

System implementation





Test for feedback & refinement

hci lab
SEOUL NATIONAL UNIVERSITY

- **Target**: not to evaluate system effectiveness but to **improve** with quick feedback

Usability test with **12 ESL learners**

- Storyfier is **more useful** than baseline without context;
- Some **have difficulties in writing** next sentence in turn-taking writing.

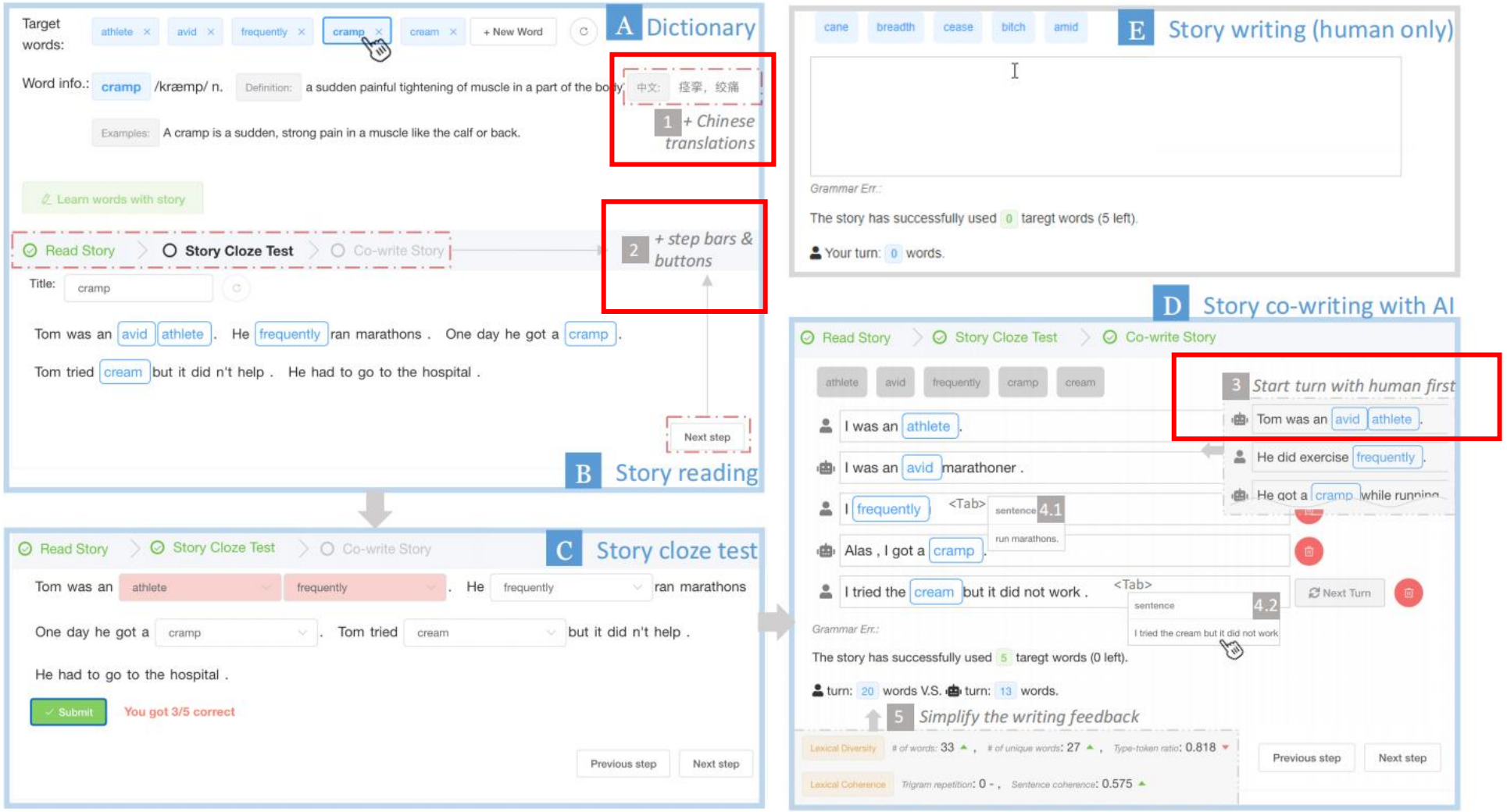
Workshop with **English teachers**

- Maximize 3 activities's value by chaining into a flow;
- Add main **translation**

Workshop with **HCI researchers**

- Use **clear widgets** in learning flow;
- **Remove metrics** to reduce distraction

System refinement

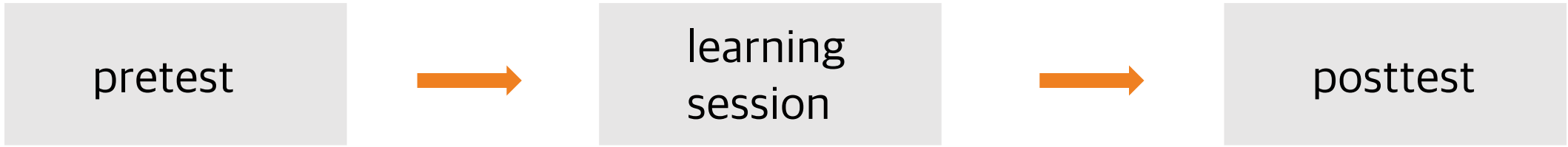


verify the impact of Storyfier on vocabulary learning

Experiment

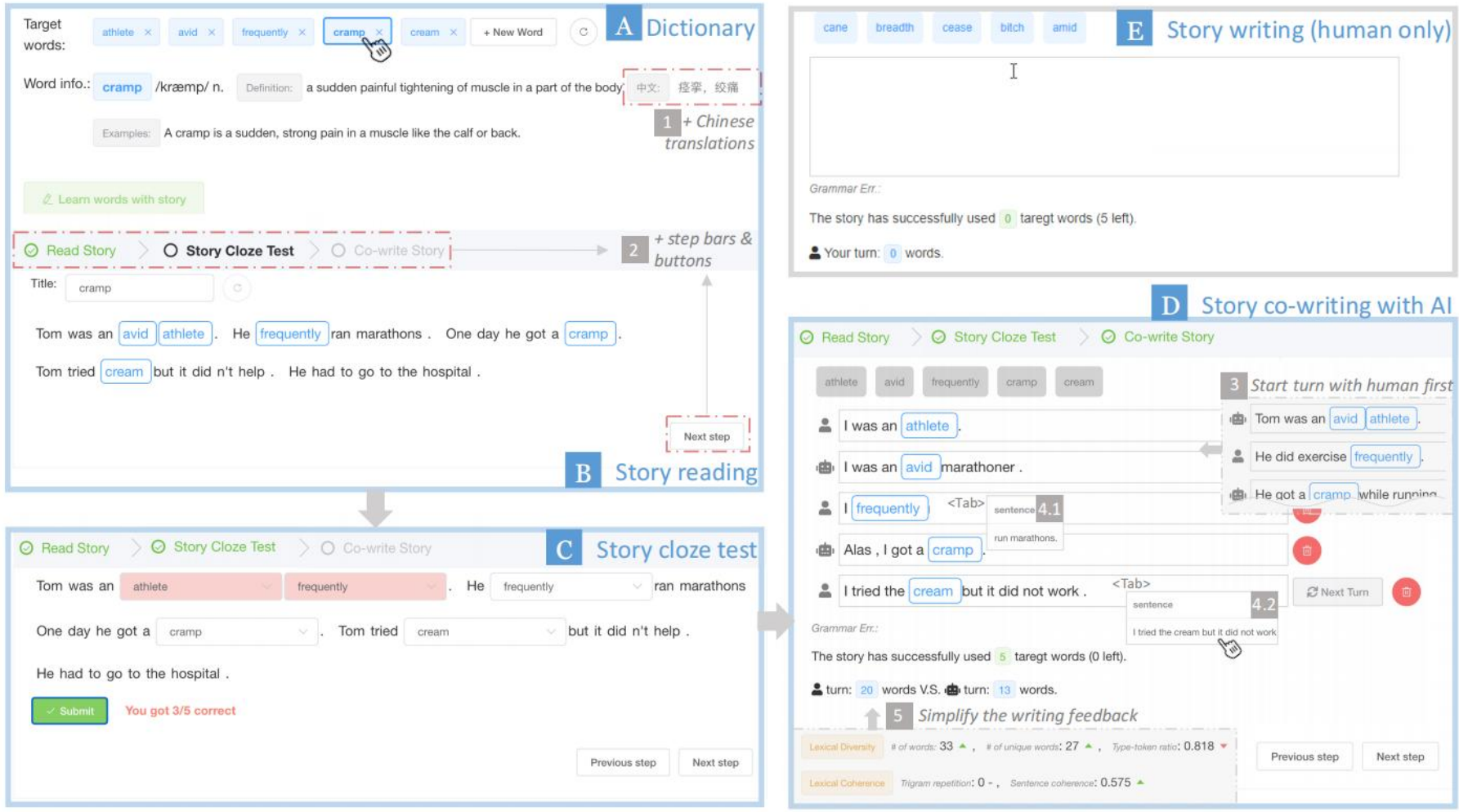
hci lab
SEOUL NATIONAL UNIVERSITY

- **Subjects**: 28 ESL Chinese students (1 / 12 have not passed CET-4)
- **Conditions**: 2 (with vs. without AI features) x 2 (read-only vs. read-cloze-write activities)
- **Research Questions**:
  - Learning outcome
  - Experience
  - Perceptions towards Storyfier
- **Procedure**:



verify the impact of Storyfier on vocabulary learning

Experiment



- Read-only
  - with AI A+B
  - without AI A
- Read-cloze-write
  - with AI A+B+C
  - without AI A+B+D

Analyses and results

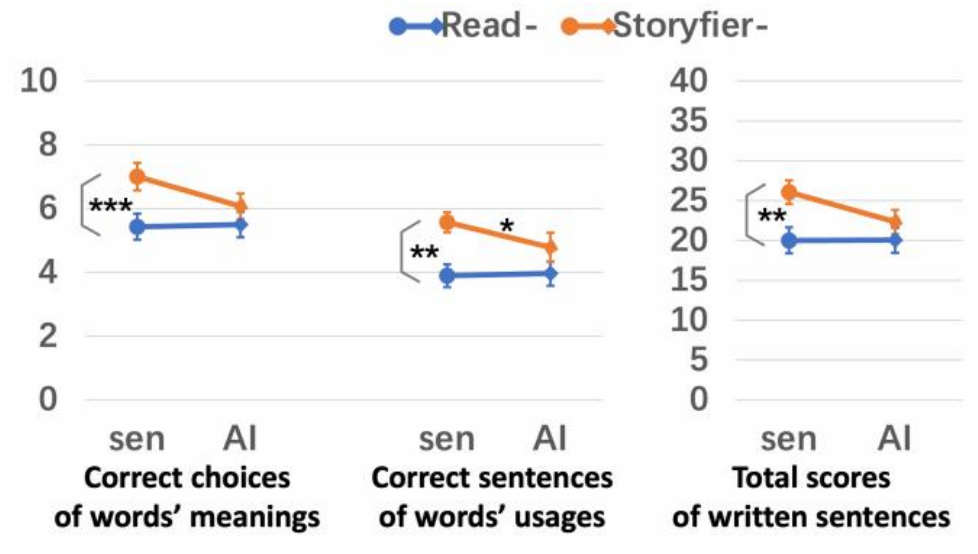


Figure 5: RQ1 results regarding numbers of correct choices on target words' meanings, numbers of sentences that correctly use target words, and total scores of the written sentences in each condition. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$.

- Read-cloze-wirte learning session has **better perfor mance** comparing to plain reading session;
- Storyfier's **AI features reduce learning gains** on ret ention words' meanings in read-cloze-wirte learning session.

Analyses and results

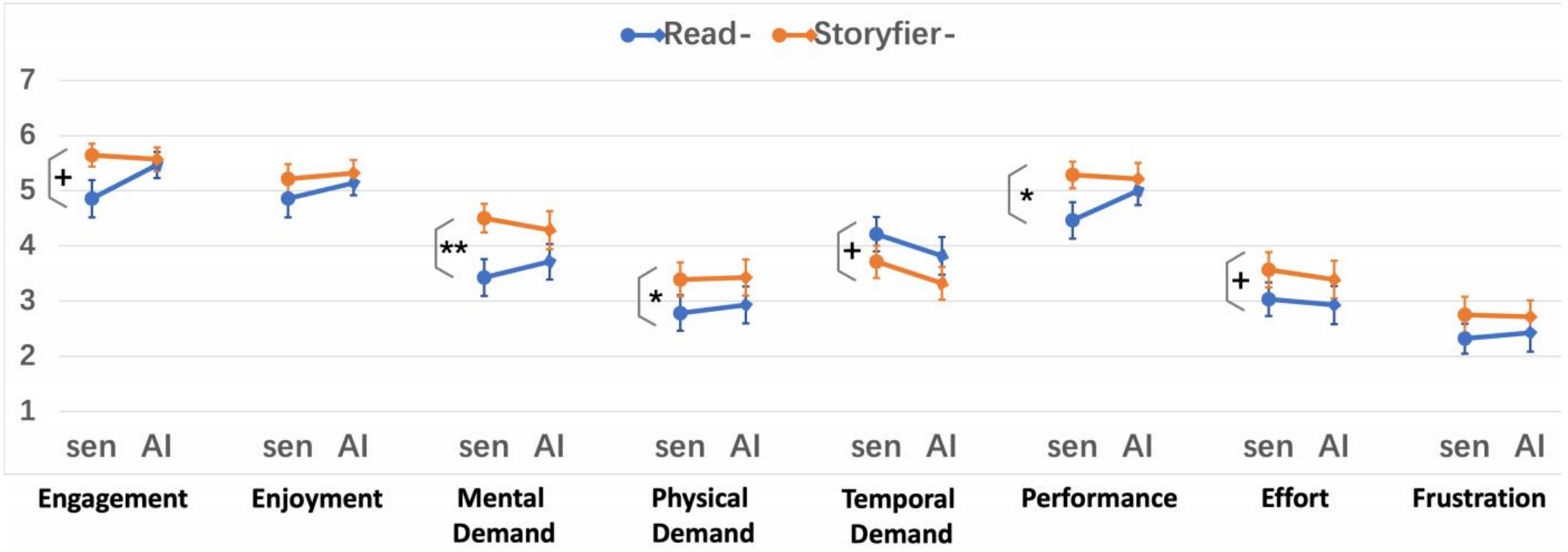


Figure 6: RQ2 results regarding perceived engagement, enjoyment, and workload in vocabulary learning sessions with Read-sen, Read-AI, Storyfier-sen, and Storyfier-AI interfaces. **: $p < 0.01$, *: $p < 0.05$, +: $p < 0.1$.

Analyses and results

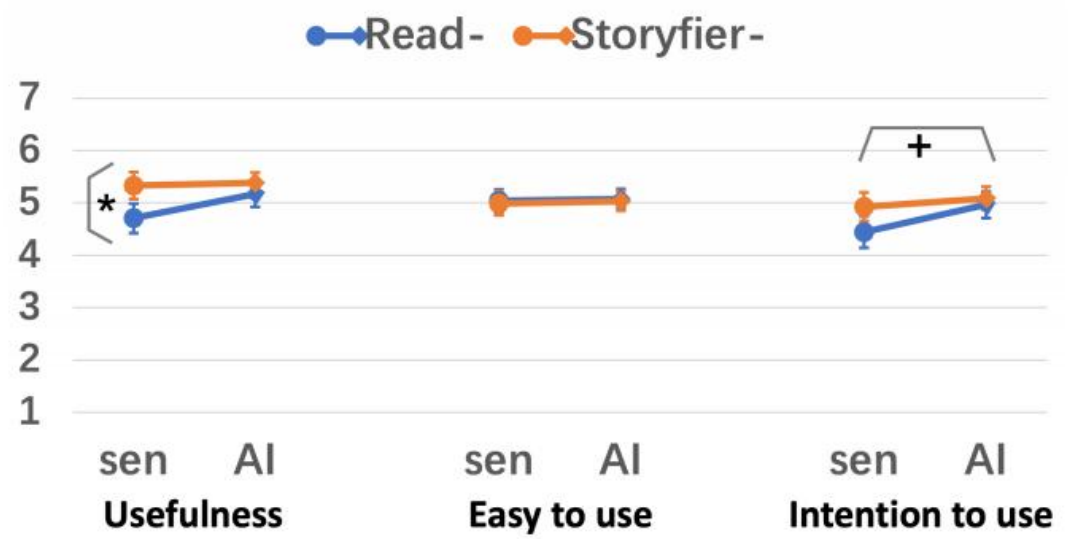


Figure 7: RQ3 results regarding user perceptions with each interface. *: $p < 0.05$, +: $p < 0.1$.

- quantitative:
  - prefer read-cloze-writing session than read-only;
  - find AI-generated stories more useful and feel ready to use in read-only session.
- qualitative:
  - participants generally prefer Storyfier with AI generative models;
  - model should be further improved regarding coherence, complexity and style of content.

Discussion

hci lab
SEOUL NATIONAL UNIVERSITY

- **generative model**:
  - **not enough coherent** when 5 target words are not naturally relevant;
  - stories contain unknown words -> content too **complex**.
- **learning activities:**
  - read-cloze-writing session provides **more learning gains**.
- **impact of genarative model in learning:**
  - though AI-generated stories do not improve learning gains -> more **coherent** content could help;
  - AI support leads to reducesd learning gains in read-cloze-writing sessions -> participants favor **assistance** fr om generative model in writing.

Discussion

hci lab
SEOUL NATIONAL UNIVERSITY

design consideration

- integrate **mutiple activities** to provide in-situ assistance;
- provide **feedback** relevant to learning goals;
- **balance** machine and human **efforts**

limitation

- Only consider **main meaning** of vocabularies;
- **quality** of stories (balance trade-off between simplicity and coherence);
- no quantitative evaluation of contribution from cloze and writing **seperately**.

hci lab
SEOUL NATIONAL UNIVERSITY

Thank you