

# Accuracy of Bilingual Chinese-speakers Using Search Systems

Zhiyan Chen<sup>1</sup> and Catherine L. Smith<sup>2</sup>

<sup>1,2</sup> Kent Kent State University, PO Box 5190 Kent, OH 44242

<sup>1</sup>zchen23@kent.edu <sup>2</sup>csmit141@kent.edu

**Abstract.** Users have substantial trust in a search engine's ability to rank results by relevance to a query. This paper reports a study that seeks to understand how English proficiency, system language (English/Chinese), and the position of relevant information on a Search Engine Result Page (SERP) affect bilingual Chinese-speakers' ability to accurately identify good items on a SERP. The results show that rank bias exists in bilingual Chinese-speakers. Results also suggest a combined effect of system language and English ability. Greater English proficiency was associated with higher accuracy on an English system, but surprisingly, the opposite was found for the Chinese system, where accuracy was lower.

**Keywords:** Bilingual native Chinese-speaker users, search engine result page (SERP), experimental studies

## 1 Introduction

By December 2016, China had 731 million internet users, 87% of whom used search engines [8]. Chinese-speaking internet users are second only to 951 million English-speaking internet users [1]. Because much quality information, ease of access, and varied uncensored opinion are available only in English, many Chinese speakers choose to use search systems in English. Two-thirds of queries submitted by Chinese students in mainland China and overseas Google using English more frequently than Chinese [2]. It is important to understand the how search behavior differs for people using search systems not in their native language.

Our research question was: For native Chinese speakers, how does a searcher's accuracy differ between Chinese and English systems when a single relevant item must be selected from a search engine result page (SERP)? To our knowledge, this question has not been addressed in prior published research.

We had three hypotheses. First was that when interacting with a system in their native language, participants would achieve higher accuracy, as compared with those using a system in English. Second, we hypothesized that searchers would be less accurate when the most relevant result for a query (the target) was displayed lower on the SERP [3]. Finally, we hypothesized that searchers with greater English proficiency would have higher accuracy as compared to those with lower proficiency.

## 2 Related Work

Related works on search behavior show that people have substantial trust in a search engine's ability to rank results. For studies of English-speaking users, 85% of the time users viewed only one page of results [6]. Users' trust in the search engine rank (rank bias) means the top two items on the SERP received more clicks and more visual attention than the rest of the items on the SERP [4]. Participant decisions were biased towards links in top positions, even for less relevant snippet abstracts [5].

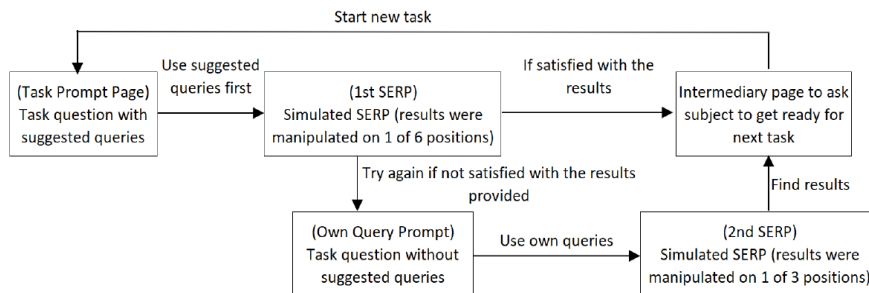
There are fewer studies of Chinese speakers' interactions with search engines. One study shows Chinese speakers have the same rank basis with results on the SERP. 85% of users browse only results retrieved on the first page [7]. However, we found no research focused on bilingual Chinese-speakers interacting with SEPRs.

## 3 Methodology

The research presented in this paper studied native Chinese speakers while they interacted with a simulated search engine in either Chinese or English, which participants used to conduct mock searches.

A total 46 subjects were recruited by a snowball method. All were native Chinese speakers who graduated from high school in mainland China and then came to study or work at a university in the midwest United States. All ranged in age from 19 to 49.

Each participant completed six different assigned informational tasks. Each task was associated with an assigned query, displayed with the task description. Participants were asked to use the assigned query first and click the "best" item they could find on the returned SERP. Participants could also submit another query of their own if they were not satisfied with the first SERP. The Figure 1 describes the procedure of the experiment:



**Fig. 1.** procedure of the experiment

The mock search engine was carefully designed to duplicate the SERP content in Chinese and English. For each task, there was only one best snippet (target). All other snippets were either partially related or completely unrelated, thus unable to provide the requested information. Targets were displayed in one of six locations, categorized

into top (1st and 2nd) and low (all other). See Table 1 for task statements and suggested queries.

**Table 1.** Search Task Statements (queries) Used in the Study.

Task	Task Question (Suggested Queries)
1	Find when the Titanic set sail for its only voyage and what port it left from. (Titanic)
2	Find the address for the Newark Airport. (Newark airport address)
3	Find out how long the Las Vegas monorail is. (Las Vegas monorail)
4	Find out the name of the building that is Piano's most famous work. (Renzo Piano)
5	Find out the size (in area) of the Oklahoma City Zoo. (Oklahoma City Zoo)
6	Find the contact number for the Sylvan Learning Center. (Sylvan Learning Center)

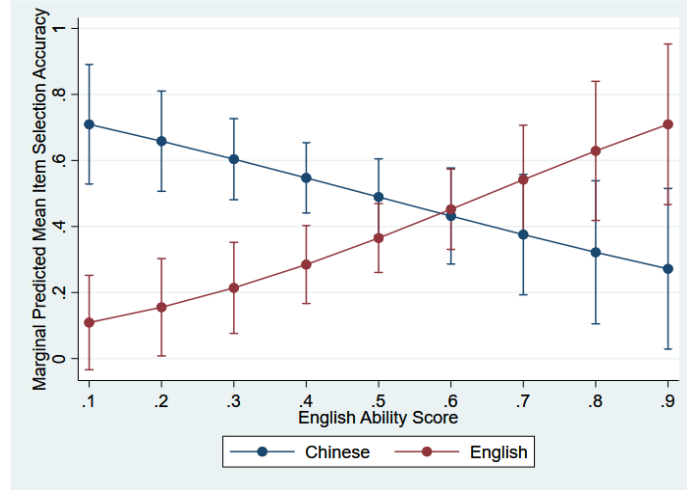
After completing all six tasks, participants also took a short English vocabulary test, which was used to control for variability in English language proficiency. English proficiency was calculated by dividing the number of correct answers by the total number of vocabulary questions (16). This ratio was used as a continuous variable.

We measured searcher accuracy for each individual subject and task as a binary variable, indicating whether the target was clicked (1) or not clicked (0).

## 4 Findings and Discussion

There were 264 valid task trials completed, including trials that used suggested queries and trials where participants used their own queries. To account for within-subject dependencies across the trials, participant was modeled as a random effect. We use a binomial logistic mixed effect model to test the results. Model fit was evaluated using Aikake information criterion (AIC). Within the 264 trials, the target was at top positions for 76, and at low positions for 188. Results show that the bilingual Chinese-speakers achieved higher accuracy when the target was located in top locations ( $p < 0.001$ ). This result is consistent with prior studies showing rank basis; subjects tend to click results displayed at the top of SERPs.

We found that English ability and system language separately did not have a significant main effect on searcher accuracy, but English ability and system language together had a significant interaction effect. Results show that the interaction of English ability and system language is significant ( $p < 0.01$ ). As hypothesized, when subjects interacted with the English system, those with greater English proficiency had higher accuracy than subjects with low proficiency. The finding is completely opposite for subjects who interacted with Chinese system. There, accuracy was lower when English proficiency was greater. The Figure 2 shows the predicted marginal effect of English proficiency on searcher accuracy for each system language. This result is surprising and worthy of further study. Discussion and speculation are limited by space constraints, however, one speculation is that English proficiency affects visual scanning patterns on the Chinese SERP. Future work using eye-tracking is planned.



**Fig. 2.** Predicted effect of interaction of system language with English ability score

1. Beitzel, S. M., Jensen, E. C., Chowdhury, A., Grossman, D., & Frieder, O. (2004, July). Hourly analysis of a very large topically categorized web query log. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 321-328). ACM. <https://doi.org/10.1145/1008992.1009048>
2. China Internet Network Information Center (2017). 39th Statistical Report on Internet Development in China. Retrieved from <http://www.ijdesign.org/ojs/index.php/IJDesign/article/view/267>.
3. Fu, T., & Karan, K. (2015). How big is the world you can explore? A study of Chinese college students' search behavior via search engines. *Procedia-Social and Behavioral Sciences*, 174, 2743-2752. <https://doi.org/10.1016/j.sbspro.2015.01.961>
4. Guan, Z., & Cutrell, E. (2007, April). An eye tracking study of the effect of target rank on web search. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 417-420). ACM. <https://doi.org/10.1145/1240624.1240691>
5. Joachims, T., Granka, L., Pan, B., Hembrooke, H., & Gay, G. (2005, August). Accurately interpreting clickthrough data as implicit feedback. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 154-161). ACM. <https://doi.org/10.1145/1076034.1076063>
6. Keane, M. T., O'Brien, M., & Smyth, B. (2008). Are people biased in their use of search engines? *Communications of the ACM*, 51(2), 49-52. <https://doi.org/10.1145/1314215.1314224>
7. Yue, Y., Patel, R., & Roehrig, H. (2010, April). Beyond position bias: Examining result attractiveness as a source of presentation bias in clickthrough data. In Proceedings of the 19th international conference on World Wide Web (pp. 1011-1018). ACM.
8. World Internet Users Statistics and 2017 World Population Stats. (2017). Retrieved November 16, 2017, from <http://www.internetworldstats.com/stats.htm>