*AN EXPERIMENTAL ANALYSIS OF THE EFFECTIVENESS AND
SUSTAINABILITY OF A CHINESE TUTORING PACKAGE*

HANG WU AND L. KEITH MILLER

UNIVERSITY OF KANSAS

This experiment evaluated the effects of training tutors to use an instructional package to teach pronunciation and translation of the Chinese language. Tutors' correct use of the package increased from 68% of trials to 92% after training, and student correct pronunciation increased from 45% to 90%, with similar effects for translation. Continued use of the package, high social validity, and extended follow-up suggest that use of the package may be sustainable.

Key words: Chinese, natural settings, sustainability, tutoring

Chinese is an increasingly important foreign language in the United States (Wu & Miller, 2007). However, Mandarin Chinese presents two difficulties to English speakers. First, the same phonemes can have entirely different meanings depending on the tonality with which they are pronounced. Second, Chinese characters do not correspond to the sounds of the words they represent. In 1979, the Chinese government addressed these difficulties by adopting the Pinyin phonetic system (Shibles, 1994). Pinyin uses Latin letters to represent the component sounds of Chinese words and diacritical marks to represent the tonality of syllables within that word. Research suggests that Pinyin improves students' ability to pronounce Chinese characters and to translate words into English (Chung, 2003). Although many American students begin their study of Chinese with Pinyin (McGinnis, 1997; Zhang, 1992), pronunciation and translation of Chinese words remain difficult challenges (Yang,

2009). Tutoring has been suggested as a solution (Yin, 2007).

Wu and Miller (2007) developed a behavior-analytic tutoring package that consisted of flash cards, Pinyin, modeling, feedback, correction, and reinforcement, to teach pronunciation of Chinese characters. A multiple baseline design across two groups of Chinese words showed that the tutoring package produced mastery pronunciation of Mandarin Chinese for one American student when implemented by the first author. The generality of the Wu and Miller (2007) tutoring package is unknown because only a single student was involved. Furthermore, it is unknown if other tutors would find the package acceptable. Therefore, the current experiment was carried out to determine (a) if the Chinese tutoring package is effective when implemented by non-behavior-analytic tutors with different students and (b) if the tutors would continue to use the package when not required to do so. In addition, we obtained social validity ratings of the procedure from both students and tutors.

METHOD

Participants and Setting

Six students and four tutors volunteered to participate. The four male and two female students were native English speakers aged 18 to 25 years. They were enrolled in a first-year Chinese language class and could write Pinyin

This article is based on a Master's thesis by the first author, under the supervision of the second author, in partial fulfillment of the requirements for the MA degree in Applied Behavioral Science at the University of Kansas. We thank committee members Claudia L. Dozier and Glen W. White.

Direct correspondence to either author at the, Department of Applied Behavioral Science, Dole Human Development Center, University of Kansas, 1000 Sunnyside Avenue, Lawrence, Kansas 66045 (e-mail: maywh@ku.edu or keithm@ku.edu).

doi: 10.1901/jaba.2012.45-217

equivalents for 200 Chinese characters. The two female and two male tutors were native Chinese speakers aged 24 to 31 years. Three were graduate students, and one was an undergraduate student. Neither the tutors nor the students were compensated for participating. None of the tutors had a background in behavior analysis. The experiment took place in a tutoring room in a midwestern university. Tutoring was not part of the students' routine coursework. Participants were informed that an unobtrusive Web camera and microphone would be used to record each session.

Measurement

Student target behavior was the percentage correct pronunciation and translation of Chinese characters during a quiz comprised of 10 Chinese words. The first author and another native Chinese speaker independently replayed recordings of student responses obtained from the Web camera for 100% of the sessions and judged the correctness of student responses. Interobserver agreement was calculated by dividing the number of agreements by the sum of agreements plus disagreements and converting the resulting number into a percentage. Agreement was 100%.

The tutor target behavior was the percentage of trials during which the tutor correctly used the tutoring package. Observers scored both a translation and a pronunciation trial each time the tutor showed the Chinese characters on one card and asked the student to translate and pronounce the characters. They scored a correct trial when the tutor correctly completed the tutoring procedure, including praising a correct answer or requiring the student to correct an error (see Table 1 for details). Data were collected on both teaching trials and review trials (see Table 1). Data on procedural fidelity were collected for 50% of the sessions. Integrity for pronunciation tutoring averaged 92% (range, 82% to 100%). Integrity for translation tutoring averaged 95% (range, 86% to 100%).

Table 1
Chinese Tutoring Steps

Step	Description
1	10 cards: Present each card by showing Chinese character (front); ask for translation and then pronunciation; when all cards are completed, go to Step 2. If correct: Praise; go to next card. If wrong: Show back of card, verbally model meaning and/or pronunciation; repeat until correct; go to next card.
2	Review 10 cards: Present cards in order taught, ask for translation and pronunciation; review all prior cards until all correct, then go to Step 3. If correct: Praise; review prior cards, repeat until all correct. If wrong: Show back of card, verbally model meaning and pronunciation; review prior cards until all correct.
3	Test 10 cards: Show Chinese character on each card; ask for translation and pronunciation; do not deliver praise or correction; when finished, go to Step 4.
4	Chat: Have a free conversation for the balance of 30-min session.

Social Validity

The researcher collected social validity data from both students and tutors. Students rated the extent to which they “would like to participate in similar tutoring programs in the future.” Tutors rated the extent to which they “would like to use this same tutoring package again.” Both groups rated their agreement on a seven-point Likert scale (1 = *extremely disagree* and 7 = *extremely agree*).

Procedure

The researcher arbitrarily selected Chinese characters for 800 words the students had not been taught and placed them on flash cards (12 cm by 8 cm). The front of each card presented Chinese characters, and the other side presented their Pinyin and English translation. Before each 30-min session, the researcher randomly selected 10 previously unused flash cards for the tutor to teach. The card sets were not assessed or equated in terms of difficulty; however, the random selection of cards created sets that were approximately equivalent over repeated sessions. Prior to the first baseline session, the researcher instructed the tutor always to quiz the student over the 10 Chinese words before the end of the session. After

finishing the quiz, the student and tutor used the remainder of the session to converse in English. Each tutor worked with one student throughout the experiment. Two tutors taught two new students during a second semester follow-up phase.

In the pretraining phase, the researcher told the tutors to use their own tutoring method to teach the students to pronounce and translate each set of 10 new words. During training, the researcher trained each tutor to use the Chinese tutoring package (Table 1; the complete tutoring protocol is available from either author in Chinese and English). The tutor read instructions in Chinese and English. The researcher modeled the procedure with the tutor playing the role of student. Next, the tutor played the tutor role while the researcher played the role of student and gave feedback. All tutors were trained until they implemented the tutoring package 100% correctly. The average training time for each tutor was 44 min (range, 35 to 48 min). During the posttraining phase, the researcher instructed the tutors to choose their preferred method to teach the students in the following sessions.

Design

We used a concurrent multiple baseline design across four tutor–student pairs. Two of four original tutors asked if they could tutor a student during the next semester. They tutored two new students as a follow-up. The researcher again told them to use whatever method they preferred.

RESULTS AND DISCUSSION

Figure 1 (left) shows the effect of training tutors how to implement the Chinese tutoring package on the performance of both students and tutors for pronunciation. For tutors, the average correct implementation of pronunciation tutoring trials increased from 68% during pretraining session to 92% during posttraining sessions. For students, the average pronun-

tion quiz score improved from 51% to 94%. Thus, training tutors to implement the Chinese tutoring package increased accurate tutor implementation and was associated with improved student pronunciation quiz scores. During follow-up, which occurred for Tutors 3 and 4, correct implementation of pronunciation tutoring trials averaged 89%, and student pronunciation quiz scores averaged 88%.

Figure 1 (right) shows a similar effect for translation of Chinese characters. For tutors, the average correct implementation of translation tutoring trials increased from 74% to 93%. For students, the average translation quiz score increased from 69% to 100%. Thus, training tutors to implement the Chinese tutoring package increased correct tutor implementation and was associated with improved student translation quiz scores. During follow-up (Tutors 3 and 4), correct implementation of translation tutoring trials averaged 93%, and student translation quiz scores averaged 97%. Overall, the tutors conducted 65% more teaching trials after they had been trained to implement the tutoring package compared with pretraining sessions (data not shown).

Tutor ratings of the extent to which they “would like to use this same tutoring package again” averaged 6.5 (of 7), and student ratings of the extent to which they “would like to participate in similar tutoring programs in the future” averaged 7. These measures suggest that the tutoring package was socially valid for both students and tutors.

The current results replicate and extend those of our previous research (Wu & Miller, 2007). These findings demonstrate generality across additional students, to tutors who do not have a background in behavior analysis, and to Chinese tutors whose use of praise and correction may be counter to their culture (Cai, 2006). Although training produced increases in tutors’ use of reinforcement from 48% to 94% and correction from 78% to 95%, it is not known how representative these

Figure 1. The left panel shows the percentage of pronunciation tutoring trials implemented correctly (bars), the percentage of correct pronunciations on session test by experimental students (diamonds), and the percentage of correct pronunciations on session test by follow-up students (circles). The right panel shows the percentage of translation tutoring trials implemented correctly (bars), the percentage of correct translations on session test by experimental students (diamonds), and the percentage of correct translations on session test by follow-up students (circles).

pretraining levels are of other potential Chinese tutors. The results of the current study also show that the tutors gave high social validity ratings to the package and used the tutoring package in the posttraining phase during which they were told to use their preferred method of tutoring; two of the tutors continued to use the package during the following semester. Furthermore, one of the tutors accepted a job teaching

Chinese at another university and reported using the procedure with her first class. Use of unobtrusive observation minimized the possibility that the tutor's continued use of the Chinese tutoring package was reactive. The current experiment replicated use of naturalistic conditions in which all researcher services and influences are minimized to approximate a normal, nonresearch situation. (e.g., DeWein

& Miller, 2008; Miller, Welsh, Altus, & Zwicker, 2006; Welsh, Miller, & Altus, 1994) and suggests that use of the tutoring package may be sustainable under nonresearch conditions.

The primary limitation of this study is the finding that tutors conducted more trials after training, which means that the effect of the tutoring package is confounded with practice. Wu and Miller (2007) found that practice did not lead to improvements without the tutoring package, but the present study did not gather data relevant to this issue. Future research could specify a set number of trials for all sessions. Another limitation is that the separate effects of the components of the tutoring package were not analyzed. Future studies might examine the necessity of praise, correction, and the Pinyin representation of the pronunciation of the Chinese characters.

The current study, along with Wu and Miller (2007), illustrates the first two steps in a strategy for developing sustainable interventions. The first step is a typical experiment in which the researcher implements the intervention to clearly establish a functional relation. The second step is to replicate the intervention with a natural implementer who is not supported by the researcher, to more closely approximate naturalistic conditions. A third step could involve creating a total package usable in the natural setting with a technology that ensures all required supports. This step has been called *institutionalization* and could lead to a sustainable intervention under the nonresearch conditions that normally characterize applied settings (e.g., Sigurdsson & Austin, 2006).

REFERENCES

- Cai, J. (2006). U.S. and Chinese teachers' cultural values of representations in mathematics education. *Mathematics Education in Different Cultural Traditions: A Comparative Study of East Asia and the West*, 9, 465–481.
- Chung, K. K. H. (2003). Effects of Pinyin and first language words in learning of Chinese characters as a second language. *Journal of Behavioral Education*, 12, 207–223.
- DeWein, M., & Miller, L. K. (2008). The effect of a teacher report on the sustainability of an intervention to facilitate engagement by a child with developmental disabilities. *Education and Treatment of Children*, 31, 333–350.
- McGinnis, S. (1997). Tonal spelling versus diacritics for teaching pronunciation of Mandarin Chinese. *The Modern Language Journal*, 81, 229–236.
- Miller, L. K., Welsh, T., Altus, D. E., & Zwicker, T. (2006). Creating sustainable behavior analytic interventions. *Current Repertoire*, 25, 3.
- Shibles, W. A. (1994). *Chinese romanization systems: IPA transliteration*. Philadelphia: University of Pennsylvania Department of Asian and Middle Eastern Studies.
- Sigurdsson, S. O., & Austin, J. (2006). Institutionalization and response maintenance in organizational behavior management. *Journal of Organizational Behavior Management*, 26, 41–77.
- Welsh, T. M., Miller, L. K., & Altus, D. E. (1994). Programming for survival: A meeting system that survives 8 years later. *Journal of Applied Behavior Analysis*, 27, 423–433.
- Wu, H., & Miller, L. K. (2007). A tutoring package to teach pronunciation of Mandarin Chinese characters. *Journal of Applied Behavior Analysis*, 40, 583–586.
- Yang, J. (2009). Chinese borrowings in English. *World Englishes*, 28, 90–106.
- Yin, J. J. (2007). Curriculum design for an entry course on Chinese characters at an American university. *Journal of Chinese Language Teachers Association*, 42, 67–75.
- Zhang, F. (1992). Native Chinese speakers and teaching Chinese as a second language. In K. Louie (Ed.), *New developments in Chinese language teaching and teacher training* (pp. 24–27). Carlton, Victoria: Asialink.

Received July 29, 2010

Final acceptance August 23, 2011

Action Editor, Einar Ingvarsson