Studia Psychologica, Vol. 68, No. 4, 2026, 345-364
https://doi.org/10.31577/sp.2026.04.953

Affective Priming in L2 among Learners Who Learn L2 Mainly in Classroom Settings: An Exploratory Study

Yang Wang, Jianlin Chen
School of Foreign Languages and Literatures, Lanzhou University, China

This study examined affective priming in English as a non-native language (L2) among learners whose L2 learning occurred mainly in classroom settings. First-year students were recruited from a university, and a set of specifically selected L2 (English) words and an evaluation task were used. In Experiment One, Group One (43 participants), who had had very limited daily use of L2, exhibited affective priming effects on the pretest, and after two weeks of movie dubbing practice, these participants exhibited similar affective priming patterns at the posttest. Before the pretest, these 43 participants were involved in the process of rating the word stimuli to ensure that the resulting norms would reflect our sample more accurately. Experiment Two was conducted to examine whether the rating of word stimuli significantly influenced these affective priming patterns. In Experiment Two, Group Two (30 participants) completed the same affective priming test used in Experiment One, and the results indicated that the process of rating the test words did not significantly influence these affective priming patterns. These findings are discussed in relation to a new model of the bilingual mental lexicon.

Key words: L2, affective priming, classroom settings, semantic and conceptual representations, bilingual mental lexicon

Introduction

In the prototypical affective priming paradigm, the affective priming effect typically results in faster evaluative judgements when prime and target words share the same valence than when they differ, particularly at a stimulus onset asynchrony (SOA) of approximately 300 ms, which typically consists of a 200-ms prime followed by a 100-ms interval (Fazio, 2001). For instance, the word “gift” as a prime can facilitate the evaluation of the target word “cake”, and “guns” facilitates “crime”. Following the seminal research of Fazio et al. (1986), affective priming has received attention from many researchers.

Correspondence concerning this article should be addressed to Yang Wang, School of Foreign Languages and Literatures, Lanzhou University, 222 South Tianshui Road, Lanzhou 730000, the People’s Republic of China. E-mail: tedwang@lzu.edu.cn
Supplementary materials are available at https://journals.savba.sk/index.php/studiapsychologica/article/view/5985/2677

Received November 5, 2025

Many studies have investigated affective priming effects across different languages, with many variants of the prototypical affective priming paradigm. For example, Hermans et al. (1998) used idiosyncratically negative and positive odors as primes for each participant, and the results revealed that target words were evaluated more quickly when preceded by affectively congruent odors. As another example, Banse (2001) used the names and faces of friends or romantic partners, and of a disliked person, as primes, and positive or negative words as targets. The results indicated that affective priming is evident for both names and faces with clearly visible primes. Additionally, Fockenberg et al. (2006) used words as both primes and targets to examine whether affective priming effects still occurred when affective primes succeeded target presentation (i.e., backward affective priming). Their results indicated that backwards affective priming "is a robust phenomenon that may reflect a rapid and continuous evaluation of environmental stimuli" (p. 799). In general, affective priming effects have "proven to be a robust and replicable phenomenon apparent in experiments using a variety of priming stimuli, target stimuli, and specific task requirements" in the first native language (L1) (Fazio, 2001, p. 117).

However, little research has focused on affective priming effects in the second language (L2) among learners who primarily learn L2 in classroom settings. The present study attempts to address this issue.

Few studies have focused on affective priming in L2. Among the few, seven relevant studies are reviewed here. In Degner et al. (2012), an evaluation task was used (with words as both primes and targets) with an SOA of 150 ms. The results revealed that "affective priming effects in L2 were only found for participants with high levels of language immersion and frequency of L2 use" (p. 181). In Tenderini et al. (2022), a music-word cross-modal affective priming paradigm was used with an SOA of 200 ms. The results revealed that "[a]mong various language background factors, only greater length of residence in the L2 context was positively related to the affective priming strength of L2 words" (p. 284). In Ahn and Jiang (2023), an emotional Stroop task was used. These authors reported that "L2 learners with higher amount of L2 use in daily life produced a significant emotion Stroop effect comparable to L1 users, while L2 learners with lower L2 use did not" (p. 1265). In Tuersun et al. (2026), an evaluation task was used (with words as both primes and targets) with an SOA of 500 ms. These authors revealed that both bilingual groups exhibited affective priming effects in L2, with stronger effects in the group reporting more frequent daily L2 use.

On the other hand, some studies have reported no affective priming effects in L2. In Experiment One of the first set of experiments in Azevedo (2016), a word-word paradigm was used in both L1 and L2 with an SOA of 300 ms. The results revealed that affective priming effects were evident only in L1, i.e., not in L2. Zhang et al. (2024) used English words as both primes and targets with an SOA of 100 ms. Although their study focused on the differences between emotion-label words and emotion-laden words in affective priming, the results revealed no affective priming effects among Chinese-English bilinguals. In Yuan and Lu (2024), words served as both primes and targets with an SOA of 300 ms. Although their study focused on the representational patterns and cognitive features of Chinese and English emotional colexification networks, the results revealed that affective priming occurred only in L1, i.e., not in L2. To the best of our knowledge, no previous study has identified affective priming effects in L2 using a word-word paradigm among learners who primarily learn L2 in classroom settings.

The first four studies suggested that affective priming in L2 occurs only when participants have frequently engaged in the daily use of L2 or have been immersed in an L2 environment. The latter three studies did not reveal affective priming in L2. However, closer consideration of the latter three studies revealed that these studies did not limit the L2 test words to those with high familiarity and/or extreme values of valence for the participants (or, at least, no such limitations were indicated).

According to the spreading activation theory (Collins & Loftus, 1975), "a [word] can be represented as a node in a [semantic] network, with properties of the [word] represented as labeled relational links from the node to other [word] nodes" (p. 408). The links or properties have "differential accessibility (i.e., strength or travel time)", and "[t]he accessibility of a property depends on how often a person thinks about or uses a property of a [word]" (p. 409), thus suggesting that the more familiar a word is to a person, the more easily the activation of that word could spread to other word nodes. Words with extreme valence (either positive or negative) could be viewed as more typical of positivity or negativity. This suggests that the typicality effect predicted by the theory (Collins & Loftus, 1975, p. 423) is stronger for such words: greater typicality enhances spreading activation within the semantic network, thereby facilitating faster affective evaluation.

For most young Chinese students in mainland China who have had very limited daily use of L2 English or have not been immersed in L2 environments, some English words have become quite familiar through many English exercises that they have completed to practice for tests (specifically, for the English test in the Gaokao, the National College Entrance Examination, which determines university admission in China). If words with extreme valence values are selected from a set of familiar words, can these words elicit affective priming effects? Furthermore, do response times for incongruent trials become shorter or longer following training?

As affective priming "is more robust in the evaluation task than in tasks (e.g., lexical decision, pronunciation) that do not depend on evaluating targets" (Storbeck & Robinson, 2004, p. 82), an evaluation task was used in the present study. It is generally assumed that affective priming takes place at the level of conceptual representation, and Pavlenko (2000a) noted that access to conceptual representation can be achieved through "a carefully coordinated investigation of contextualized language use (such as elicited language production or role play)" (p. 2). Therefore, an updated version of roleplaying, namely, movie dubbing practices, was used as a training method to develop or improve affective priming in L2.

Accordingly, the research questions (RQs) investigated in the present study are as follows:

RQ1: Do Chinese–English bilinguals exhibit affective priming in an English word–word evaluation paradigm in the pretest or posttest when they do not frequently use their L2 in daily life and have not been immersed in an L2 environment?

RQ2: How do response times (RTs) for congruent and incongruent trials change from the pretest to the posttest?

RQ3: How does accuracy (ACC) change from the pretest to the posttest?

Participants were included in the process of rating the word stimuli before the pretest to ensure that the resulting norms more accurately reflected our sample. Therefore, another experiment was conducted to examine whether the process of rating word stimuli significantly influenced affective priming patterns. RQ4 is thus as follows: Does rating the

test words significantly influence affective priming patterns?

Overall, two experiments were designed and conducted with an English word-word evaluation paradigm. Experiment One aimed to answer the first three RQs, whereas Experiment Two aimed to answer the final RQ.

Experiment One

Participants

Two hundred thirty-six first-year students recruited from a university in mainland China participated in the subjective rating of 243 English words (their participation in the rating task was rewarded with additional course credits). These individuals were non-English majors from various departments or schools, such as the School of Physics, the School of Chemistry, and the Department of Medicine. They were all students in B-level English classes (in both experiments — Experiment One and Experiment Two — all participants, whether involved in the rating process or in the affective priming test, were first-year students in B-level English classes). Traditionally, at this university, at the beginning of the first term, all first-year students other than those from the School of Foreign Languages and Literatures take the English Replacement Test (ERT) to be placed in different levels of English classes (from a set of four levels, i.e., A, B, C, and D, with A as the lowest level and D as the highest).

Among the 236 first-year students, 45 students volunteered to participate in Experiment One (their participation in Experiment One was rewarded with gifts). Data of two students were excluded from subsequent analyses because one was left-handed and the other was trilingual. Among the remaining 43 students, 16 were men and 27 were women. All 43 students had normal or adjusted-to-normal vision, had no history of mental or physiological illness, and were right-handed. The descriptive statistics obtained through the Language Background and Proficiency Survey, which was adapted from Gong et al. (2024) (see Supplementary Material 1) and which the 43 participants (Group One) completed after the pretest for Experiment One, are presented in Table 1.

Table 1 clearly shows that the participants had very limited daily use of L2 English, indicating that the level of L2 immersion among the participants was relatively low. In other words, these individuals had primarily learned L2 English in classroom settings. Notably, the survey asked how many hours the participants had spent on English outside regular class hours per week **in recent months** (i.e., in the several months following their enrolment at this university). During the period between the participants' completion of the Gaokao and their enrolment at this university, most of these individuals did not use English at all, which was confirmed by subsequent interviews with some of the students. Before these students completed the Gaokao, most of them were too busy preparing for the Gaokao to spend time communicating in English, watching English movies/TV, playing English video games or reading English novels (middle-school English education in China is generally Gaokao oriented, and the English test included in the Gaokao covers only listening comprehension, reading comprehension, cloze for grammar/language use, and writing), which was once again confirmed by the subsequent interviews.

Materials

The English words used for the affective priming test were selected from the lines of actors in the movie *The Shawshank Redemption.* The selection was made by reference to a list of

Table 1 *Mean (M) and standard deviation (SD) of language background and self-ratings of the 43 participants*

Variable	*M* (*SD*)
Age (years)	18.21 (0.51)
Age of English acquisition (years)	8.65 (2.70)
Hours per week spent on English outside regular class hours in recent months apart from preparation and review for English class	
Communication in English	2.27 (2.35)
Listening to English songs	5.76 (8.16)
Watching English movies and TV	1.64 (1.62)
Playing English video games	0.57 (1.61)
Reading English materials	1.13 (1.28)
Listening to English materials	2.58 (2.09)
Self-Rated Proficiency in English	
Listening	3.49 (1.97)
Speaking	4.21 (1.78)
Reading	4.93 (1.26)
Writing	4.14 (1.32)

Note. Self-ratings of proficiency were scored on a scale ranging from 0 (no proficiency) to 10 (high proficiency).

13915 English lemmas with norms pertaining to valence, arousal, and dominance on a 9-point Likert scale (1–9), which was provided by Warriner et al. (2013) (only negative words with a valence value smaller than or equal to 3 and the positive words with a valence value equal to or greater than 7 were selected from the list) and the word list provided in the Chinese government document *The English Curriculum Standards for Senior High School* (2020), which, practically speaking, serves as a minimum-required list of English words that senior high school students must practice in English exercises to prepare for the English test in the Gaokao. From the actors' lines, 143 words (including 65 positive words and 78 negative words) were selected. An additional 100 English words (including 50 positive words and 50 negative words) for the practice procedure were selected by reference to both of the sources mentioned above. The 236 students rated these 243 words online. The rating job took place during one of the author's English classes. This author served as both a supervisor and an expositor for the rating survey.

The survey covered three dimensions (see Supplementary Material 2 for the online survey): familiarity, valence, and arousal. The corresponding items were scored on a 7-point Likert scale (1–7), where 1 indicated extremely unfamiliar, extremely unhappy, and extremely calm, whereas 7 indicated extremely familiar, extremely happy, and extremely excited. When the value of familiarity was 4 or lower, the corresponding data for valence and arousal were deleted. Based on the students' ratings, 72 words (including 36 positive words

and 36 negative words) from 23 movie clips were selected with regard to the three dimensions, i.e., valence, familiarity, and arousal, as well as word length. These 72 words were then semi-randomized into 36 pairs: 9 negative-negative pairs, 9 positive-positive pairs, 9 negative-positive pairs, and 9 positive-negative pairs (see Supplementary Material 4). The parts of speech of the prime and the target were not manipulated because the meta-analysis of affective priming studies by Herring et al. (2013) found that this factor does not significantly influence affective priming effects. Among the 72 words, 4 were not covered by the word list provided by the Chinese government document *The English Curriculum Standards for Senior High School* (2020): "shit", "grant", "insane", and "fucking"; additionally, "grant" and other words with inflections were not found in the word (lemma) list provided by Warriner et al. (2013). However, the participants might already have become familiar with words not covered by these two resources in various ways (including through practice for the Gaokao).

All statistical analyses for the present study were performed in R (Version 4.5.2; R Core Team, 2025). The word stimulus validation analyses were conducted at the word level. Practical equivalence was assessed using the predefined smallest effect sizes of interest (SESOIs) of ±1 letter for word length and ±1 scale unit for familiarity and arousal.

The positive and negative words differed strongly in terms of valence, with a mean difference (*MD* = positive − negative) of 3.85, 95% confidence interval (CI) of [3.62, 4.07], confirming the intended manipulation, $t(67.93) = 34.23$, $p < .001$, Cohen's $d = 8.07$. Word length exhibited no significant differences, $t(66.46) = 0.91$, $p = .364$, $d = 0.22$, with an *MD* of 0.36 below the SESOI, although equivalence was not formally established according to the two one-sided tests (TOST) procedure ($p_{lower} < .001$, $p_{upper} = .055$); the 90% CI [-0.30, 1.02] slightly exceeded the upper equivalence bound by 0.02 letters. Familiarity did not exhibit significant differences, $t(64.27) = 1.37$, $p = .176$, $d = 0.32$, and equivalence was established (TOST $p_{lower} < .001$, $p_{upper} < .001$), with the 90% CI [-0.05, 0.49] falling entirely within the SESOI. Arousal exhibited slight differences, as negative words were rated more arousing than positive words, $t(47.71) = -2.09$, $p = .042$, $d = -0.49$, but equivalence was established (TOST $p_{lower} < .001$, $p_{upper} < .001$), with the 90% CI [-0.42, -0.05] falling within the SESOI. Overall, these results demonstrate a robust valence manipulation, while the two word groups were broadly comparable with the other stimulus characteristics. Descriptive statistics of the two groups of words are presented in Table 2.

Table 2 clearly indicates that the participants were generally quite familiar with the 72 words (negative words: 6.32; positive words: 6.54) and that these words exhibited quite extreme values of valence (negative words: 2.47; positive words: 6.32). Considering the possibility of fatigue on the part of

Table 2 *Descriptive statistics for word-level variables by valence category*

Variable	Positive *M* (*SD*)	Negative *M* (*SD*)	*MD*
Length	6.17 (1.86)	5.81 (1.47)	0.36
Valence	6.32 (0.52)	2.47 (0.43)	3.85
Arousal	4.86 (0.62)	5.10 (0.27)	-0.23
Familiarity	6.54 (0.58)	6.32 (0.79)	0.22

participants, only 60 words were used in the practice procedure (30 positive and 30 negative words were semi-randomized into 7 negative–negative pairs, 7 positive–positive pairs, 8 negative–positive pairs, and 8 positive–negative pairs). These thirty word pairs were used for the practice procedure in both Experiments One and Two.

Procedure

Pretest

All 45 participants took the pretest (i.e., the affective priming test using the 36 pairs of English words) before the training (i.e., the movie dubbing practice), 8 days after the rating process. The test was designed using PsychoPy v2024.1.5 software and launched on a Lenovo laptop. All RT tests in the present study were launched on this laptop. After the participants provided written consent, each individual completed the test in a quiet laboratory while wearing over-ear headphones; during this time, one of the authors stood by in case the participant needed help during the practice procedure. At the beginning of the practice and the test, the participants read the written instructions in Chinese on the screen, which required the participants to view two words that were presented consecutively (in short succession) carefully and then to judge the second word by pressing the key "b" with the right index finger to indicate "negative" or pressing the key "g" with the left index finger to indicate "positive" as accurately and quickly as possible (the same instruction was also delivered orally prior to the experiment). The difference between the practice procedure and the test procedure was that, in the practice procedure, feedback was provided to indicate whether participants pressed the wrong key or not, whereas in the test procedure, no such feedback was provided.

In the test procedure, trials were randomized for each participant via PsychoPy software. In each trial, a fixation cross ("+") was shown at the center of the screen for 1000 ms to center participants' attention; this image was accompanied by a soft beeping sound during the first 700 ms. Subsequently, the English word prime was displayed for 200 ms at the same location, followed by a 100-ms blank interval. An English target word then appeared at the same location and remained on the screen until the participant responded. After a response was provided, a 2000-ms blank screen was shown to allow participants to rest their eyes before the next trial began. The SOA was fixed at 300 ms. Figure 1 illustrates a single-trial demonstration.

After all 45 participants completed the pretest outside regular class hours (in two days), they were provided with 23 video clips from the movie *The Shawshank Redemption* containing the 72 words included in the test and the corresponding scripts (see Supplementary Material 3); they were then required to complete the Language History and Proficiency Survey on the Teaching and Learning Platform of the university (assistance was provided by one of the authors when the participants experienced problems while completing the survey).

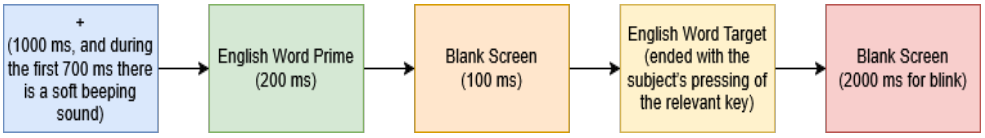


Figure 1 Single-trial demonstration.

Training Stage

Following the procedures above, the participants could use the 23 video clips to practice independently (on their smartphones, iPads, or PCs) or go to a multimedia classroom equipped with a laser projector, a large screen, and loudspeakers at a time convenient to them (the participants were encouraged to focus on the actors' emotions during practice). To motivate the participants, they were informed that a movie-dubbing competition would be held and that the top three-fourths of participants would receive a 64 GB USB 3.0 flash drive from a well-known brand as a reward. Thirty-seven participants indicated that they wished to participate in the movie dubbing competition.

Three days after the pretest, ten classroom activities were conducted. A social media group was established that included one of the authors and the 43 participants with the aim of facilitating immediate communication. At convenient times for the participants, one of the authors visited the classroom to conduct practice sessions. The first session involved viewing the entire film *The Shawshank Redemption*, which was attended by approximately 40 participants. The remaining nine sessions were used to practice movie dubbing based on the 23 video clips, either between participants and one of the authors or among participants themselves. The average attendance across these nine sessions was approximately 15 participants. This level of attendance was understandable because all the classroom activities were held outside regular class hours, and the schoolwork faced by first-year students during their first term was relatively heavy, including many different courses, assignments, and laboratory work. Across these nine sessions, the 23 clips were practiced one by one in chronological order. During each session, participants in attendance were randomly selected by one of the authors to ensure that each participant practiced on two or three different clips. At the beginning of the session, the volume of the clips was set high and gradually reduced thereafter. Participants practiced by imitating the actors' speech and emotions. Each session lasted approximately two hours. The two-week practices began on Sunday of the first week and ended on Saturday of the second week. Since all of the participants were scheduled to take the College English Test Band 4 (CET-4), which was important to them (in this university, undergraduate students who are not majoring in English must pass the CET-4 to obtain their bachelor's degrees), on the Saturday morning of the third week, the posttest was not held immediately after this two-week practice period. We considered that the participants might not have had sufficient time (since all of the tests were conducted outside regular class hours) or ability to focus on the posttest while they were busy preparing for the CET-4.

Posttest

On the Saturday afternoon of the third week, the posttest (involving the same affective priming test as in the pretest) began and was completed over two days, with two participants absent (one was left-handed and the other was trilingual). In other words, 43 participants completed the posttest. Two weeks after the posttest, the movie dubbing competition was held, with 17 participants (this participation rate was understandable because the students were busy preparing for their final examinations in all the courses they had participated during the first term, one week later). Finally, all 17 participants were rewarded with a 64 GB USB 3.0 flash drive from a well-known brand.

Results and Discussion

No participants showed an overall ACC below 60% at either time point. ACC analyses included all 3,096 trials. For RT analyses, incorrect responses (319 trials), trials with RTs longer than 3000 ms or shorter than 200 ms (18 trials), and RTs exceeding ±2.5 *SD*s from the participant × time × congruency mean (70 trials) were excluded. The descriptive statistics are presented in Table 3.

Log-transformed RTs were analyzed using a linear mixed-effects model (LMM) with fixed effects of Time, Congruency, and their interaction, as well as participant-level random intercepts and by-participant random slopes for Time and Congruency. ACC was analyzed using a binomial generalized linear mixed-effects model (GLMM) with the same fixed effects structure and by-participant random slopes for Congruency.

The LMM revealed a significant main effect of Congruency, $F(1, 40.84) = 75.18$, $p < .001$, standardized $\beta = -0.14$, indicating overall slower responses for incongruent trials than for congruent trials. The main effect of Time was not significant, $F(1, 41.72) = 1.72$, $p = .197$, standardized $\beta = 0.05$, indicating that RTs did not improve overall from the pretest to the posttest. The Time × Congruency interaction was also not significant, $F(1, 2575.43) = 0.49$, $p = .485$, standardized $\beta = 0.01$, indicating that the strength of the affective priming effect did not change significantly from the pretest to the posttest.

Planned pairwise comparisons based on the estimated marginal means (EMMs) derived from the log-transformed model were conducted to examine the priming effects at each time point. At the pretest, the participants responded significantly more slowly to incongruent trials than to congruent trials (log-scale difference = -0.079, 95% CI [-0.11, -0.05], $t(158) = -5.69$, $p < .001$, $d = -0.32$), corresponding to a raw difference of 73 ms (972 vs. 899 ms). At the posttest, participants once again responded significantly more slowly to incongruent trials than to congruent trials (log-scale difference = -0.093, 95% CI [-0.12, -0.06], $t(155) = -6.72$, $p < .001$, $d = -0.37$), corresponding to a raw difference of 86 ms (938 ms vs. 852 ms).

Additional planned comparisons tested the effect of Time within each trial type. For congruent trials, RTs did not change significantly from the pretest to the posttest, log-scale difference = 0.040, 95% CI [-0.014, 0.094], $t(54.4) = 1.48$, $p = .145$, $d = 0.16$. For incongruent trials, RTs also did not change significantly from the pretest to the posttest, log-scale difference = 0.027, 95% CI [-0.028, 0.081], $t(55.8) = 0.97$, $p = .335$, $d = 0.11$.

Regarding the GLMM, the main effect of Congruency was significant, $\chi^2(1) = 5.92$, $p = .015$, $b = 0.18$, indicating that the overall ACC was significantly lower for incongruent trials than for congruent trials. The main effect of Time was not significant, $\chi^2(1) =$

Table 3 *Descriptive statistics for RTs and ACC across time and congruency*

Time	Congruency	Mean RT (ms) (*SD*)	RT Trials (*n*)	Mean ACC (%) (*SD*)	ACC Trials (*n*)
Pre	Congruent	899 (335)	690	92.0 (27.2)	774
Pre	Incongruent	972 (338)	646	86.7 (34.0)	774
Post	Congruent	852 (269)	686	91.2 (28.3)	774
Post	Incongruent	938 (300)	667	88.9 (31.4)	774

0.21, p = .649, b = -0.03, indicating the absence of any overall changes in ACC from the pretest to the posttest. The Time × Congruency interaction was also not significant, $\chi^2(1) = 1.68$, p = .194, b = 0.08. Overall, the ACC remained stable from the pretest to the posttest, and no significant Time or interaction effects were observed.

Now, the first three RQs can be addressed. With respect to RQ1 (Do Chinese–English bilinguals exhibit affective priming in an English word–word evaluation paradigm in the pretest or posttest when they do not frequently use their L2 in daily life and have not been immersed in an L2 environment?), the results were as follows: In Chinese–English bilinguals who had little daily use of English or no L2 immersion, both the RT and ACC data revealed a significant main effect of Congruency. Namely, incongruent trials elicited slower RTs and lower ACCs than congruent trials did. Robust affective priming effects were observed in RTs at both the pretest and the posttest. These effects were observed for the specifically selected English words, which were characterized by high familiarity and extreme valence and were validated prior to the experiment. Regarding RQ2 (How do RTs for congruent and incongruent trials change from the pretest to the posttest?), the results were as follows: The RTs for both congruent and incongruent trials were not significantly faster at the posttest than at the pretest. Regarding RQ3 (How does ACC change from the pretest to the posttest?), the results were as follows: The GLMM revealed no significant main effect of Time and no significant Time × Congruency interaction, indicating that overall ACC did not change from the pretest to the posttest.

To examine possible valence asymmetries, Time (pre vs. post), Congruency (congruent vs. incongruent), and Target Valence (negative vs. positive) were entered as within-participants factors in both the RT and ACC mixed-effects models. For the valence asymmetry RT analyses, trimming was conducted within each participant × time × congruency × target valence cell (4 trials were excluded in this step) to preserve the factorial structure of the additional valence factor. After cleaning, 2,755 RT trials remained for analysis. Descriptive statistics for these analyses are presented in Table 4.

In the RT analysis, there was a significant main effect of Congruency, $F(1, 1309.74) = 70.98$, $p < .001$, standardized β = -0.14, indicating slower responses for incongruent trials than for congruent trials. There was also a significant main effect of Target Valence, $F(1, 2663.97) = 4.78$, p = .029, standardized β = 0.04, reflecting slightly faster responses

Table 4 *Descriptive statistics for RTs and ACC across time, congruency, and target valence*

Time	Congruency	Target Valence	Mean RT (ms) (*SD*)	RT Trials (*n*)	Mean ACC (%) (*SD*)	ACC Trials (*n*)
Pre	Congruent	Negative	936 (357)	342	89.4 (30.8)	387
Pre	Congruent	Positive	900 (374)	361	94.6 (22.7)	387
Pre	Incongruent	Negative	1016 (382)	324	84.0 (36.7)	387
Pre	Incongruent	Positive	982 (389)	337	89.4 (30.8)	387
Post	Congruent	Negative	872 (293)	348	90.2 (29.8)	387
Post	Congruent	Positive	874 (319)	357	92.2 (26.8)	387
Post	Incongruent	Negative	961 (324)	338	87.9 (32.7)	387
Post	Incongruent	Positive	965 (362)	348	89.9 (30.1)	387

to positive targets than to negative targets overall. Critically, however, the Congruency × Target Valence interaction was not significant, $F(1, 2664.58) = 0.07$, $p = .788$, standardized $\beta = 0.004$, indicating that the size of the congruency effect did not reliably differ between positive and negative targets. The Time × Congruency × Target Valence interaction was likewise not significant, $F(1, 2664.56) = 0.08$, $p = .783$, standardized $\beta = 0.004$, suggesting that training did not differentially influence priming across valence categories.

The GLMM analysis of ACC revealed a similar overall pattern. There was a significant main effect of Congruency, $\chi^2(1) = 6.30$, $p = .012$, $b = 0.19$, indicating higher accuracy for congruent trials than for incongruent trials. There was also a significant main effect of Target Valence, $\chi^2(1) = 11.87$, $p < .001$, $b = -0.21$, reflecting higher accuracy for positive targets than for negative targets. Importantly, the Congruency × Target Valence interaction was not significant, $\chi^2(1) = 0.34$, $p = .563$, $b = -0.04$, indicating that the congruency effect on accuracy did not differ between positive and negative targets. The Time × Congruency × Target Valence interaction was also not significant, $\chi^2(1) = 0.15$, $p = .694$, $b = -0.02$.

Overall, these findings indicate that affective priming operated similarly for positive and negative target words. Although positive targets elicited slightly faster and more accurate responses overall in terms of RTs and ACC, there was no evidence that the priming effect itself was stronger for one valence category than for the other or that training differentially altered priming across valence categories.

In Experiment One of the first set of experiments of Azevedo (2016), analyses were conducted at the participant level following absolute RT trimming (specifically, trials with RTs below 300 ms and above 2500 ms were excluded). Participants with an overall ACC below the level of chance (.50) were removed from the analyses. No *SD*-based trimming procedure was applied. Following the analytical approach of Azevedo, the data from Experiment One in the present study yielded the following results: for RTs, a significant main effect of Congruency was observed, $F(1, 42) = 63.71$, $MSE = 3882$, $p < .001$, $\eta_p^2 = .60$; the main effect of Time was not significant, $F(1, 42) = 2.65$, $MSE = 26378$, $p = .111$, $\eta_p^2 = .06$; the interaction between Time and Congruency was not significant, $F(1, 42) = 0.77$, $MSE = 2514$, $p = .385$, $\eta_p^2 = .018$; paired-samples t tests revealed significant affective priming effects at both time points (pretest: $t(42) = 5.25$, $p < .001$, $d = 0.80$; posttest: $t(42) = 7.41$, $p < .001$, $d = 1.13$).

Therefore, the differences between the current results and those of Azevedo (2016) are unlikely to be due to using different analytical approaches. Two factors may account for this discrepancy. First, in the present study, the age of acquisition (AOA) was 8.65 (*SD* = 2.70), whereas in Azevedo (2016), "most of them [started to learn L2] when they were 10 years old" (p. 68). Second, the word stimuli used in that context were not specifically selected as they were in the present study. Zhang et al. (2024) and Yuan and Lu (2024) did not select word stimuli in the same manner as in the present study.

Notably, the 43 participants in the present study had rated relevant English words 8 days before the pretest. It was thus necessary to examine whether completing the rating task could significantly influence the observed affective priming patterns. Therefore, Experiment Two was conducted.

Experiment Two

Participants

An additional 37 first-year students who had not participated in the rating task were will-

ing to participate in Experiment Two. First, these individuals were asked to complete the Language Background and Proficiency Survey (the same survey used in Experiment One). Based on their survey data, 30 students (14 men and 16 women) were selected and paid 80 RMB (approximately 11 US dollars) for their work in Experiment Two.

The ERT, administered to first-year students at the beginning of their first term, is provided by the Foreign Language Teaching and Research Press, a leading English textbook publisher in China. This test includes listening, reading, translation, and writing sections, with a maximum score of 100 points. The listening and reading sections are scored by an automated system, whereas the translation and writing sections are graded by experienced English language teachers. Therefore, the test has acceptable validity and reliability.

The differences between Group One and the 30 participants in Experiment Two (Group Two) in the 12 English-related variables listed in Table 1, as well as the total ERT score, were examined at the participant level. Practical equivalence was evaluated with the predefined SESOIs: ±0.5 years for age and age of English acquisition, ±1 hour for weekly English exposure, ±1 scale unit for self-rated proficiency, and ±5 points for the total ERT score. The descriptive statistics for the 13 variables across the two groups are presented in Table 5.

MD, Welch *t* tests*, d,* SESOIs, and equivalence tests are summarized in Table 6.

None of the variables exhibited statistically significant differences between groups, with the exception of weekly hours spent on communication in English, $t(70.3) = 2.32$, $p = .023$, $d = 0.531$, with an *MD* of 1.04 hours only slightly exceeding the predefined SESOI by 0.04 hours. Equivalence was not estab-

Table 5 *Descriptive statistics of the 13 English-related variables between the two groups*

Variable	Group 1 (*n* = 43) Mean (*SD*)	Group 2 (*n* = 30) Mean (*SD*)	*MD*
Age (years)	18.21 (0.51)	18.17 (0.59)	0.04
Age of English acquisition (years)	8.65 (2.70)	8.33 (2.01)	0.32
Hours per week spent on English outside regular class hours in recent months apart from preparation and review for English class			
Communication in English	2.27 (2.35)	1.23 (1.47)	1.04
Listening to English songs	5.76 (8.16)	4.77 (6.03)	0.99
Watching English movies and TV	1.64 (1.62)	2.23 (2.29)	−0.59
Playing English video games	0.57 (1.61)	0.83 (1.53)	−0.26
Reading English materials	1.13 (1.28)	0.64 (0.98)	0.49
Listening to English materials	2.58 (2.09)	2.35 (1.61)	0.23
Self-Rated Proficiency in English			
Listening	3.49 (1.97)	3.63 (1.47)	−0.14
Speaking	4.21 (1.78)	3.73 (1.82)	0.48
Reading	4.93 (1.26)	5.13 (1.55)	−0.20
Writing	4.14 (1.32)	4.27 (1.17)	−0.13
The ERT			
Total score on the ERT	57.15 (1.14)	57.20 (1.07)	−0.05

Note. *MD* = Group one minus group two.

Table 6 *Group differences in the English-related variables (MD, t, d, and equivalence tests)*

Variable	*MD*	*t*(df)	*p*	*d*	90% CI	SESOI	TOST *ps*
Age (years)	0.04	0.32 (56.7)	0.751	0.077	[-0.181, 0.266]	±0.5	<.001, <.001
Age of English acquisition (years)	0.32	0.58 (70.7)	0.567	0.134	[-0.600, 1.236]	±0.5	.071, .371
Hours per week spent on English outside regular class hours in recent months apart from preparation and review for English class							
Communication in English	1.04	2.32 (70.3)	0.023	0.531	[0.294, 1.787]	±1	<.001, .536
Listening to English songs	0.99	0.60 (70.7)	0.552	0.138	[-1.77, 3.760]	±1	.117, .498
Watching English movies and TV	-0.59	-1.22 (48.7)	0.229	-0.298	[-1.407, 0.223]	±1	.203, <.001
Playing English video games	-0.26	-0.71 (64.3)	0.481	-0.168	[-0.884, 0.357]	±1	.026, <.001
Reading English materials	0.49	1.83 (70.3)	0.071	0.426	[0.044, 0.927]	±1	<.001, .028
Listening to English materials	0.23	0.53 (70.3)	0.598	0.123	[-0.493, 0.952]	±1	.003, .040
Self-Rated Proficiency in English							
Listening	-0.14	-0.36 (70.6)	0.72	-0.083	[-0.817, 0.527]	±1	.019, .003
Speaking	0.48	1.11 (61.7)	0.271	0.265	[-0.240, 1.192]	±1	<.001, .113
Reading	-0.20	-0.59 (54.1)	0.555	-0.144	[-0.775, 0.369]	±1	.012, <.001
Writing	-0.13	-0.43 (66.9)	0.667	-0.102	[-0.617, 0.363]	±1	.002, <.001
The ERT							
Total score on the ERT	-0.05	-0.19 (65.0)	0.847	-0.046	[-0.487, 0.386]	±5	<.001, <.001

Note. 90% CI = 90% confidence interval for *MD*; TOST *ps* = *p* values for lower and upper in equivalence tests.

lished for this variable (TOST p_{lower} < .001, p_{upper} = .536). Eight out of the total 13 variables exhibited equivalence: Age, Playing English video games, Reading English materials, Listening to English materials, Self-rated proficiency in Listening, Reading, and Writing, and total score on the ERT. For these variables, both TOST p_{lower} and p_{upper} were < .05, and the 90% CIs were entirely within the SESOIs. For the remaining four nonequivalent variables, the *MD*s were numerically smaller than the SESOIs, and the results of the Welch *t* tests did not indicate significant differences (*p*s ≥ .229). In summary, although some variables did not meet the requirements of formal equivalence, the groups were largely comparable in practice. Any minor differences observed are unlikely to introduce systematic bias into subsequent analyses.

Materials and Procedure

The materials used for the affective priming test (including the practice word pairs and test word pairs) were the same as those in Experiment One, and the test procedure, with the exception that Group Two completed the test within less than one day and one week after Group One finished the posttest.

Analyses, Results, and Discussion

No participants in Group One at the pretest or in Group Two at the test showed an overall ACC below 60%. For RT analyses, incorrect responses (335 trials), trials with RTs longer than 3000 ms or shorter than 200 ms (24 trials), and RTs exceeding ±2.5 *SD*s from the group × participant × congruency mean (51 trials) were excluded, leaving 2,218 trials.

However, at the participant level, the *SD* of the trimmed mean RT was 202.06 ms, and the pooled within-group SD was 203.38 ms. Using the latter estimate and the current sample sizes (n_1 = 43, n_2 = 30; total N = 73), the resulting 90% CI half-width for the difference between group means was ±80.63 ms. This level of imprecision substantially exceeds commonly

used equivalence bounds in cognitive psychology. Consequently, regarding the present sample size, equivalence cannot be confirmed for RTs under either a ±30 ms or ±50 ms SESOI, regardless of how small the observed mean differences are. Power analyses further indicated that, given the observed participant-level variability, very large samples would be required to establish RT equivalence. Specifically, approximately 788 participants per group would be required for the TOST procedure within a ±30 ms SESOI with 80% power, and approximately 285 participants per group would be required for a ±50 ms. Therefore, equivalence testing was not considered informative for RTs given the substantial imprecision associated with the present sample size.

Log-transformed RTs and ACC were analyzed using mixed-effects models following the analytical approach used in Experiment One. In this between-groups analysis, Group (Group One pretest vs. Group Two) replaced Time as a fixed effect interacting with Congruency, and the models included participant-level random intercepts and by-participant random slopes for Congruency. The descriptive statistics are presented in Table 7.

The LMM (logRT) revealed a significant main effect of Congruency, $F(1, 2003.15) = 57.02$, $p < .001$, standardized $\beta = -0.13$, indicating overall slower responses for incongruent trials than for congruent trials. The main effect of Group was not significant, $F(1, 70.72) = 0.13$, $p = .725$, standardized $\beta = 0.03$, suggesting no overall RT differences between the groups. The Group × Congruency interaction was also not significant, $F(1, 2003.15) = 0.52$, $p = .472$, standardized $\beta = 0.01$, indicating that the magnitude of the priming effect did not differ significantly between the groups.

Planned comparisons based on EMMs confirmed significant priming effects within each group. In Group One, the participants responded significantly more slowly to incongruent trials than to congruent trials (log-scale difference = -0.0786, 95% CI [-0.11, -0.05], $t(67.2) = -5.42$, $p < .001$, $d = -0.30$), corresponding to a raw difference of 73 ms (972 vs. 899 ms). In Group Two, the participants also responded significantly more slowly to incongruent trials than to congruent trials (log-scale difference = -0.0951, 95% CI [-0.13, -0.06], $t(75.3) = -5.32$, $p < .001$, $d = -0.36$), corresponding to a raw difference of 103 ms (982 vs. 879 ms). The difference in priming magnitude between groups was not significant (log-scale difference = 0.017, [-0.03, 0.06], $t(71.9) = 0.72$, $p = .474$, $d = 0.06$).

Additional planned comparisons were used to test the group differences within each congruency condition. RTs did not differ significantly between the groups for congruent trials (log-scale difference = 0.025, 95% CI [-0.073, 0.124], $t(71.1) = 0.52$, $p = .608$, $d = 0.10$) or for incongruent trials (log-scale difference = 0.009, 95% CI [-0.091, 0.109], $t(71.2) = 0.18$, $p = .861$, $d = 0.03$). Collapsing across congruency, there was no overall RT difference

Table 7 *Descriptive statistics for RT and ACC by group and congruency*

Group	Congruency	Mean RT (ms) (*SD*)	RT Trials (*n*)	Mean ACC (%) (*SD*)	ACC Trials (*n*)
G1(pre)	Congruent	899 (335)	690	92.0 (27.2)	774
G1(pre)	Incongruent	972 (338)	646	86.7 (34.0)	774
G2	Congruent	879 (326)	471	89.3 (31.0)	540
G2	Incongruent	982 (401)	411	79.3 (40.6)	540

between the groups (log-scale difference = 0.017, 95% CI [-0.079, 0.113], $t(71.1) = 0.35$, $p = .725$, $d = 0.06$).

The GLMM revealed a significant main effect of Congruency, $\chi^2(1) = 17.18$, $p < .001$, $b = 0.28$, indicating that the overall ACC was lower for incongruent trials than for congruent trials. A significant main effect of Group was observed, $\chi^2(1) = 8.19$, $p = .004$, $b = 0.23$, indicating a higher overall ACC in Group One than in Group Two. The Group × Congruency interaction was not significant, $\chi^2(1) = 0.88$, $p = .347$, $b = -0.06$, indicating no significant differences in priming effects between groups.

Follow-up between-group comparisons revealed no significant differences in accuracy for congruent trials, odds ratio (*OR)* = 1.39, 95% CI [0.94, 2.06], $z = 1.66$, $p = .098$, $d = 0.18$, but significantly greater accuracy was observed in Group One than in Group Two for incongruent trials, $OR = 1.79$, 95% CI [1.16, 2.75], $z = 2.66$, $p = .008$, $d = 0.32$, likely reflecting differences in testing conditions between the groups: Group Two completed the test within less than one day, whereas Group One completed the pretest in two days.

Planned comparisons revealed significant accuracy priming effects in both groups. In Group One, the participants were significantly more likely to respond correctly on congruent trials than on incongruent trials, $OR = 1.54$, 95% CI [1.07, 2.21], $z = 2.32$, $p = .020$, $d = 0.24$. In Group Two, a similar effect was observed, $OR = 1.98$, 95% CI [1.35, 2.90], $z = 3.51$, $p < .001$, $d = 0.38$.

In summary, the overall performances of Group One (pretest) and Group Two were generally comparable in affective priming. In particular, Group Two reproduced the priming patterns observed in Group One (pretest), i.e., significant RT priming and ACC priming, with no significant RT differences in either congruent or incongruent trials.

Group One participated in the process of rating the words (specifically covering the test words) before the pretest. Group Two did not engage in the process of rating the words before the test. First, affective priming occurs at the automatic level (Fazio, 2001), and it is difficult to imagine that the rating task (which involved rating one word at a time within seconds) could significantly influence affective priming patterns. Second, two weeks of movie-dubbing practice did not produce a substantial change in the priming patterns from pretest to posttest for Group One. Therefore, the most likely explanation for these findings is that the rating task did not significantly influence the affective priming patterns. This result answers RQ4 (Does rating the test words significantly influence affective priming patterns?): Rating the test words did not significantly influence affective priming patterns. Consequently, the pretest of Group One is replicable in terms of the affective priming patterns.

A New Model of the Bilingual Mental Lexicon to Account for the Pretest

Mechanisms Underlying Affective Priming, the Distinction between Conceptual and Semantic Representations (Knowledge), and Affective Priming at Conceptual and Semantic Levels

Before the new model is presented, it is necessary to make the following introductions. There are two popular accounts of affective priming effects: the spread activation mechanism (SAM) and the response competition mechanism (RCM). According to the SAM (Klauer & Musch, 2003), a prime, when it is perceived, activates its own node within a semantic network. This activation subsequently propagates to nodes of targets whose evaluative valence matches that of the prime more quickly than to nodes of targets with mismatched valence. Consequently, whenever

the prime and target share the same evaluative consistency, the target is processed more quickly. According to the RCM (Fazio, 2001), the prime activates an attitude, which in turn prepares the individual to respond in a certain way. If the target is congruent with the prime in valence, the response is facilitated because the response pathway is already partially activated. In contrast, if the target is incongruent with the prime in valence, then the response pathway initially activated by the prime must be inhibited to provide an accurate response to the target, and the result is a slow response because of such interference.

Semantic knowledge refers to "definitions of words, as well as explicitly stated grammatical rules and conventions of linguistic usage" (Pavlenko, 2000b, p. 33). Notably, native speakers could still explain the definitions of words, as well as the related grammatical rules and conventions of the linguistic usage of L1, either with or without the technical terms invented by grammarians. Conceptual knowledge is "experientially acquired [non-declarative] conceptual representations" (Pavlenko, 2000b, p. 33) and covers elements such as "imagery, schemas, motor programs, auditory, tactile, and somatosensory representations" (Pavlenko, 2000a, p. 2).

Logically, "the attitude activated as a consequence of the prime", as described in the RCM by Fazio (2001, p. 120), should not exist in a semantic network (the semantic level). Therefore, the conceptual representations correspond to the RCM, whereas the semantic representations correspond to the SAM. With the ERP method, Eder et al. (2012) investigated the contributions of the RCM and SAM to affective priming in L1. Based on previous studies, these authors used the LRP as an online marker of RCM priming and the N400 as an online marker of SAM priming. The results revealed that both "semantic priming and response priming constitute affective priming effects" (p. 436) in the evaluative task. These findings suggest that in L1 affective priming, two kinds of distinct priming are operative: semantic priming through the SAM at the semantic level and response priming through the RCM at the conceptual level. With the distinction between semantic and conceptual representations in L2, it is reasonable to assume that this pattern also exists in L2 affective priming.

The New Model that Can Account for the Pretest

In the pretest for Experiment One, the 43 participants who had previously used L2 (English) daily only to a very limited extent, or had not been immersed in an L2 environment, exhibited affective priming effects. It is possible that most of the test words had reached the conceptual level through the enormous number of English exercises these individuals completed for the Gaokao; however, no research has focused on this topic, and relevant theoretical support remains lacking. It is more likely that most of the words had reached the level of semantic representations but not that of conceptual representations because 1) these individuals had previously experienced little L2 socialization (see Table 1) and 2) they had been aware of the definitions of these words (68 out of the 72 words were covered by the minimum-required word list provided by the Chinese government document *The English Curriculum Standards for Senior High School,* 2020) and were quite familiar with these words (for information concerning word familiarity, see Table 2) as a result of the English exercises that they had completed rather than through extensive L2 socialization. Therefore, it could be concluded that the affective priming of the 43 participants in the pretest of Experiment One occurred at the semantic level through the SAM.

L1 is usually acquired through socialization in emotion-rich contexts, thus allowing L1 to reach the conceptual and semantic levels. L2 can also reach the conceptual level through L2 socialization. This is consistent with the findings of Degner et al. (2012), Tenderini et al. (2022), Ahn and Jiang (2023), and Tuersun et al. (2026), who reported that affective priming effects occurred in L2 only when the participants had engaged in frequent daily use of L2 or have been immersed in L2 environments, thus indicating that most of the test words had reached the conceptual level as well as the semantic level for these participants through socialization.

Following an analysis of the advantages and disadvantages of the previous models of the bilingual mental lexicon, Pavlenko, an advocate for the distinction between semantic and conceptual knowledge, proposed a model (Pavlenko, 2009, p. 147) (Figure 2).

However, despite its many advantages, the model cannot explain the fact that some L2 words have reached the semantic level but have not yet reached the conceptual level since, in Pavlenko's model, semantic representations are defined as "links between words and concepts, as well as words and other words" (p. 148).

Based on Pavlenko's (2009) model, a new model (Figure 3) is developed to explain the findings of the present study. (Since the cross-language transfer issues between different levels and between different elements in the mental lexicon are too complex and lie far beyond the scope of the present study, they are not demonstrated in this new model.) In this model, most of the test words had reached the semantic level (as indicated by the solid lines) for the 43 participants in the present study, but had not yet reached the conceptual level (as indicated by the dotted lines). In Degner et al. (2012), Tenderini et al. (2022), Ahn and Jiang (2023), and Tuersun et al. (2026), most of the test words had reached the conceptual level as well as the

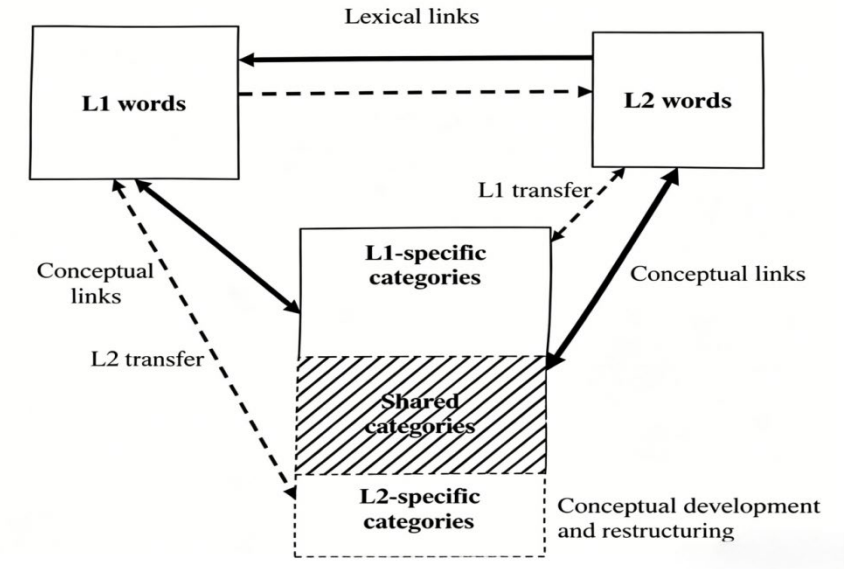


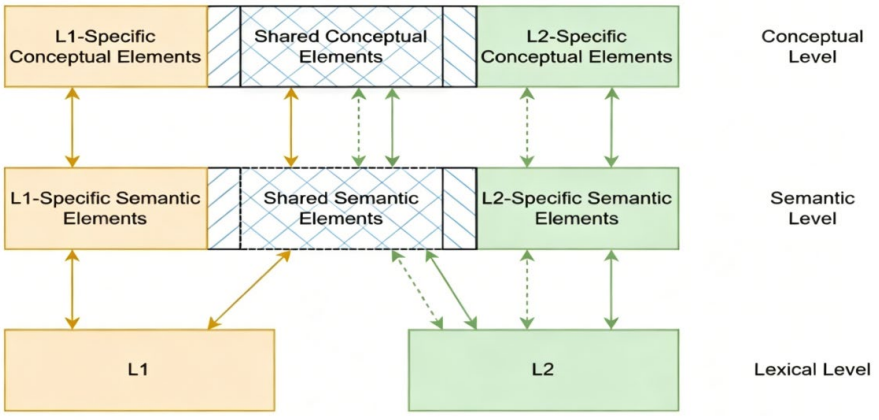
Figure 2 Pavlenko's model of the bilingual mental lexicon.

semantic level through L2 socialization (as indicated by the solid lines between the lexical level and semantic level and between the semantic level and the conceptual level) for the participants. In Azevedo (2016), Zhang et al. (2024), and Yuan and Lu (2024), most of the test words had not yet reached the semantic level (as indicated by the dotted lines between the lexical and semantic levels) for the participants. Regarding L1, in general, words in the participants' native language have reached both the conceptual and semantic levels. Therefore, there are only solid lines between the lexical level and the semantic level and between the semantic level and the conceptual level.

Since the present new model is largely based on Pavlenko's model, it is necessary to clarify the differences between these two models. The first difference is that Pavlenko defines semantic knowledge as "the mapping between words and concepts" and "connections between words" (2009, p. 148). In contrast, the present model maintains that semantic knowledge "encompasses definitions of words, as well as explicitly stated grammatical rules and conventions of linguistic usage" (Pavlenko, 2000b, p. 33). The second difference lies in Pavlenko's definition of L2 learning: "[L2] learning is seen as a gradual process, taking place in implicit memory" (2009, p. 150). The new model posits that L2 learning occurs in both explicit and implicit memory.

Conclusion

For the present study, first-year students (from a university in mainland China) who learned L2 (English) mainly in classroom settings were recruited to investigate affective priming in their L2. English words with extreme valence values were selected from a set of familiar words, and an evaluation task (involving 36 prime-target pairs where the target was to be judged as positive or negative) was used. Two experiments were conducted. The results showed that there were affective priming effects in L2 among learn-



Note. In the model, the lexical level contains the forms of words with their "phonological and morphosyntactic properties" (Pavlenko, 2000a, p. 2).

Figure 3 A new model of the bilingual mental lexicon.

ers who had had very limited daily use of L2 English. A new model of the bilingual mental lexicon has been proposed to account for the findings of the present study.

However, the present study has obvious limitations. First, during the study, we regretfully did not notice that the CET-4 time overlapped with our training sessions. In future studies, irrelevant factors such as the CET-4 could be avoided. Second, due to limited access to specialized equipment, the present study did not incorporate physiological or neural indicators such as the skin conductance response, event-related potential, or functional magnetic resonance imaging. The results of the present study are thus limited to the behavioral level and lack cross-validation with physiological and neural data; nevertheless, this limitation highlights a direction for future studies. Third, the present study used only an evaluation task; other task types (such as a lexical decision task or a pronunciation task) should be included in future studies.

Acknowledgments

We would like to thank all participants in the study for dedicating their valuable time to our research. This study was supported by National Social Science Fund of China (No. 23BYY142).

Authors' ORCID
Yang Wang
https://orcid.org/0000-0002-6570-7774
Jianlin Chen
https://orcid.org/0000-0002-7357-3622

References

Ahn, S., & Jiang, N. (2023). Can adult learners sense L2 emotional words automatically? The role of L2 use on the emotional Stroop effect. *Second Language Research*, *39*(4), 1265–1278. https://doi.org/10.1177/02676583221131256

Azevedo, C. M. M. e. (2016). *Social cognitive consequences of differences in the emotional grounding of concepts: The role of embodiment* (ISBN: 9781083605771)[Doctoral dissertation, ISCTE – Instituto Universitario de Lisboa (Portugal)]. ProQuest Dissertations & Theses. https://www.proquest.com/dissertations-theses/social-cognitive-consequences-differences/docview/2033065305/se-2?accountid=11988

Banse, R. (2001). Affective priming with liked and disliked persons: Prime visibility determines congruency and incongruency effects. *Cognition and Emotion, 15*(4), 501–520.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Degner, J., Doycheva, C., & Wentura, D. (2012). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition, 15*, 181–189. https://doi.org/10.1017/S1366728911000095

Eder, A. B., Leuthold, H., Rothermund, K., & Schweinberger, S. R. (2012). Automatic response activation in sequential affective priming: An ERP study. *Social Cognitive and Affective Neuroscience, 7*(4), 436–445. https://doi.org/10.1093/scan/nsr033

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion, 15*(2), 115–141. https://doi.org/10.1080/0269993004200024

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*(2), 229–238.

Forkenberg, D. A., Koole, S. L., & Semin, G. R. (2006). Backward affective priming: Even when the prime is late, people still evaluate. *Journal of Experimental Social Psychology, 42*, 799–806. https://doi.org/10.1016/j.jesp.2005.12.003

Gong, Y., Hansen, K., & Chen, J. (2024). The (dis)engagement of different components of inhibitory control in trilingual language control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 50(7)*, 1095–1111. https://doi.org/10.1037/xlm0001315

Hermans, D., Baeyens, F., & Eelen, P. (1998). Odours as affective-processing context for word

evaluation: A case of cross-modal affective priming. *Cognition and Emotion, 12*(4), 601–613.

Herring, D. R., White, K. R., Jabeen, L. N., Hinojos, M., Terrazas, G., Reyes, S. M., Taylor, J. H., & Crites, S. L., Jr. (2013). On the automatic activation of attitudes: A quarter century of evaluative priming research. *Psychological Bulletin, 139*(5), 1062–1089.

Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7–49). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Ministry of Education of the People's Republic of China. (2020). *The English Curriculum Standards for Senior High School.* http://www.moe.gov.cn/srcsite/A26/s8001/202006/W020200603315372317586.zip

Pavlenko, A. (2000a). New approaches to concepts in bilingual memory. *Bilingualism: Language and Cognition, 3*(1), 1–4.

Pavlenko, A. (2000b). What is in a concept? *Bilingualism: Language and Cognition, 3*(1), 31–36.

Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. In A. Pavlenko (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches* (pp. 125–160). Clevedon, UK: Multilingual Matters.

Storbeck J., & Robinson M. D. (2004). Preferences and inferences in encoding visual objects: A systematic comparison of semantic and affective priming. *Personality and Social Psychology Bulletin, 30,* 81–93. https://doi.org/10.1177/0146167203258855

Tenderini, M. S., de Leeuw, E., Eilola, T. M., & Pearce, M. T. (2022). Reduced cross-modal affective priming in the L2 of late bilinguals depends on L2 exposure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(2), 284–303. https://doi.org/10.1037/xlm0000889

Tuersun, A., Feng, Y., & Guo, J. (2026). Reciprocal role of semantic meaning in affective processing in L2: Evidence from Uyghur-Mandarin bilinguals. *Journal of Psycholinguistic Research*, *55*, 9. https://doi.org/10.1007/s10936-025-10194-w

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Yuan, L., & Lu, S. (2024). A priming study on the processing mechanism of emotional colexification in Chinese and L2 English from the perspective of typology. *Modern Foreign Languages*, *47*(3), 406–418. https://doi.org/10.20071/j.cnki.xdwy.20240328.011

Zhang, J., Wu, C., Meng, Y., & Yuan, Z. (2024). An ERP study on affective priming of second language: The emotion word type effect in unmasked and masked priming paradigms. *Advances in Cognitive Psychology, 20*(1), 1–11. https://doi.org/10.5709/acp-0408-y