Title

The Association between Humor Comprehension and Subjective Social Well-being in Non-native English Speakers

Authors

Xu, Haoyin
Nakanishi, Masaki
Coulson, Seana

The Association between Humor Comprehension and Subjective Social Well-being in Non-native English Speakers

Haoyin Xu (hyx002@ucsd.edu) and Masaki Nakanishi (masaki@sccn.ucsd.edu)
Swartz Center for Computational Neuroscience, University of California San Diego
9500 Gilman Drive, La Jolla, CA 92093, USA

Seana Coulson (scoulson@ucsd.edu)
Department of Cognitive Science, University of California San Diego
9500 Gilman Drive, La Jolla, CA 92093, USA

Abstract

The goal of language learning should be to fit in with the language community, and this often requires much more than linguistic knowledge. Although both social wellness in a second language (L2) society and L2 humor comprehension require sophisticated social and cultural knowledge beyond linguistic proficiency, their direct association has not previously been tested. Here we developed a novel method to assess different stages of humor comprehension (i.e., detection and appreciation) and conducted a series of experiments to explore its relationship with subjective social well-being in non-native English speakers. The results revealed significant correlations between language anxiety and social connectedness with both humor detection and humor appreciation in the L2. The findings suggest that the ability of L2 humor detection can be a hallmark of pragmatic proficiency and social wellness in an L2 community.

Keywords: Foreign Language Anxiety; Humor Comprehension; Social Connectedness

Introduction

In recent decades, the use of humor in teaching a second language (L2) has received substantial attention as an effective pedagogical approach (Tarone, 2000; N. D. Bell, 2005). L2 teachers have been encouraged to use language play including humor in a classroom because it motivates L2 learners and facilitates their learning process. Recent studies have suggested that the use of humor in a classroom also enhances students' pragmatic skills since it requires complex processing of lexical items and creative uses of linguistic resources (Tarone, 2000; N. D. Bell, 2005, 2009). Although grammatical proficiency itself has not been shown to be a good predictor of pragmatic competence, L2 humor skills could be a marker of comprehensive language proficiency (N. D. Bell, 2005). However, even with sophisticated linguistic competence, the detection and comprehension of verbal humor may still be challenging for non-native speakers (NNSs) (Nelms, 2001).

To properly understand humor, one must be able to detect incongruous elements that deviate from his/her prediction about incoming word(s) and to resolve the incongruity (Shultz, 1972; Ritchie, 1999). This ability requires mastery of sociocultural norms that go beyond essential linguistic proficiency. In fact, failed humor in cross-cultural conversation often occurs due to differences in ethnic norms of communication (N. Bell, 2007). In addition, it has been shown that L2 speakers need to reach a certain level of pragmatic and sociocultural competence in their L2 to be able to appropriately appreciate verbal humor (X. Chen & Dewaele, 2019).

To access sociocultural proficiency, Chen *et al.* found evidence showing that cultural intelligence is positively related to social connectedness, defined as a sense of social connection to the new community (A. Chen, Lin, & Yan, 2021). Given the significance of sociocultural knowledge in L2 humor competence, this suggest that for L2 speaker, the social connectedness could also be a predictive measurement. However, despite the commonality between social wellness in an L2 community and L2 humor competence, i.e. both require sophisticated linguistic and sociocultural proficiency, little study has empirically investigated their direct association.

To address this gap in the literature, the present study examines the relationship between L2 humor comprehension, psycholinguistic factors, and subjective social well-being. To do so, we conducted a series of experiments. In the first experiment, we prepared joke stimuli and validated them based on the judgement and ratings of native English speakers (NSs). In the second and third experiment, we used these validated joke stimuli to assess the ability of a group of non-native English speakers (NNSs) to detect (Experiment 2) and appreciate (Experiment 3) jokes. Moreover, to probe the relationship between social well-being and joke comprehension, we also assessed NNS using two scales validated in previous studies: 1) foreign language anxiety scale (Horwitz, Horwitz, & Cope, 1986) and 2) social connectedness scale (Lee, Draper, & Lee, 2001). Given their shared reliance on linguistic and sociocultural competence, we hypothesized joke comprehension ability would be significantly correlated with measures of social well-being.

Experiment 1: Stimulus Validation

The purpose of this experiment was to prepare and validate a set of joke stimuli for the humor detection and appreciation studies described in Experiment 2 and Experiment 3, respectively. Accordingly, we created 56 pairs of stimuli – one a joke harvested from social media and the other an edited version intended to serve as a control stimulus. We asked American-born native English speakers to complete the humor detection and appreciation tasks we planned to administer to L2 English speakers in Experiments 2 and 3. The goal was thus to verify that our joke materials were considered jokes by native English speakers and that the control version of each stimulus was not. Similarly, we asked native speakers

3398

to rate how humorous our materials were. This was done, first to ensure that each joke was considered funnier than its control, and, second, to serve as a baseline for comparison with non-native English learners' ratings for the same sentences.

Participants

We recruited 179 undergraduate students as subjects in this experiment. An additional 19 people were enrolled in the study but were excluded because they failed attention checks described below in the Procedure section. All subjects were at least 18 years old and were native English speakers who were born and raised in the United States. Subjects were divided into two groups: 84 (59 females, mean age = 20.30 ± 1.84) were assigned to group A and asked to complete the humor detection task described below, and 95 (78 females, mean age = 20.73 ± 2.71) were assigned to group B and assigned to the humor appreciation task described below.

Materials

A total of 56 English puns were obtained from the website of Indian Hills Community Sign[1] and they were treated as joke stimuli. From these we prepared 56 non-joke versions by replacing word(s) that would make the sentence more serious or neutral. For example, the joke sentence "What did Noah see at night? Flood lights." had as a non-funny 'Statement' counterpart "What did Noah see at night? Portable Lights." For each pair of stimuli, subjects saw only a single (randomly chosen) version – either the joke or the non-joke version of the stimulus. The pseudo-random assignment of versions was such that each subject saw approximately 28 jokes and 28 non-jokes over the course of the study.

Procedure

The experiment was conducted on a cloud-based survey platform via Qualtrics online survey software. The survey began with a consent form approved by the Human Research Protections Program of the University of California, San Diego. After granting informed consent, participants advanced to either the humor detection task (Group A) or the humor appreciation task (Group B). In both tasks, one version of each of the 56 experimental sentences (either joke or non-joke) was presented at a time and participants were either asked to classify it as a joke or a non-joke (Group A) or to rate the funniness of the sentence (Group B) using a 5-point Likert scale ranging from 1: not funny at all, to 5: extremely funny. In order to prevent subjects from adopting a strategy in which they clicked the same response on every trial, we interspersed "catch" trials that explicitly asked participants to respond with a particular response. For example, participants in Group A were told "Please select joke" or "Please select statement" on catch trials; participants in Group B were explicitly told (for example) "Please select 3". Any participant who did not respond correctly to at least one of these requests

---

[1]https://www.facebook.com/IndianHillsCommunitySign/

was excluded from data analysis. This resulted in the exclusion of 19 participants out of an initial enrollment of 198. Results below reflect data from a total of 179 subjects.

Analysis: Stimulus Validation

Each of the stimulus pairs was evaluated by its detection accuracy and subjective rating. In humor detection, the accuracy of each stimulus pair was calculated as the number of subjects who answered correctly divided by the total number of subjects (i.e., 84). Each stimulus pair was also statistically evaluated by a chi-squared test to assess the null hypothesis that there was no difference in the distribution of subjects' responses to its joke and non-joke versions.

In joke appreciation, non-parametric permutation tests with unpaired t-tests, in which the number of permutations was set to 1,000, were used to test the null hypothesis that there was no difference in the distributions of the subjective ratings between joke and non-joke stimuli for each stimulus pair.

Results

In the joke detection task, 50 stimulus pairs both achieved an accuracy score above 50 % and reached statistical significance on the chi-squared test ($p < 0.05$). In the joke appreciation task, 51 stimulus pairs conformed to the expected pattern in which the joke version was rated significantly funnier that its control ($p < 0.05$). A total of 43 stimulus pairs met the criteria in both joke detection and appreciation, and thus were chosen for use in Experiments 2 and 3. The average detection accuracy across the 46 stimuli selected in the validation was 75.78 ± 11.06. The median ratings (interquartile range: IQR) on the joke stimuli were 2.45 (1.68 - 3.17) while those for the non-joke stimuli were 0.50 (0.27 - 0.86); jokes were rated as significantly more funny than the non-jokes ($t = 13.60$, $p < 0.001$).

Experiment 2: Humor detection

This experiment aims to examine the relationship between the accuracy of humor detection in non-native English speakers (NNS) and their anxiety regarding communication in L2, their frequency of language use, and their subjective social well-being in the U.S..

Participants

Following the exclusion of several participants who failed to answer correctly on "catch" trials, there were 143 NNS participants in the humor detection task (110 females, mean age = 20.78 ± 2.31). Each subject was compensated with academic course credits. All subjects were at least 18 years old and provided informed consent to participate in the experiment.

Data Collection

The experimental procedure was adapted from experiment 1. The NNS subjects were asked to perform the humor detection task described in Experiment 1, in which they are asked
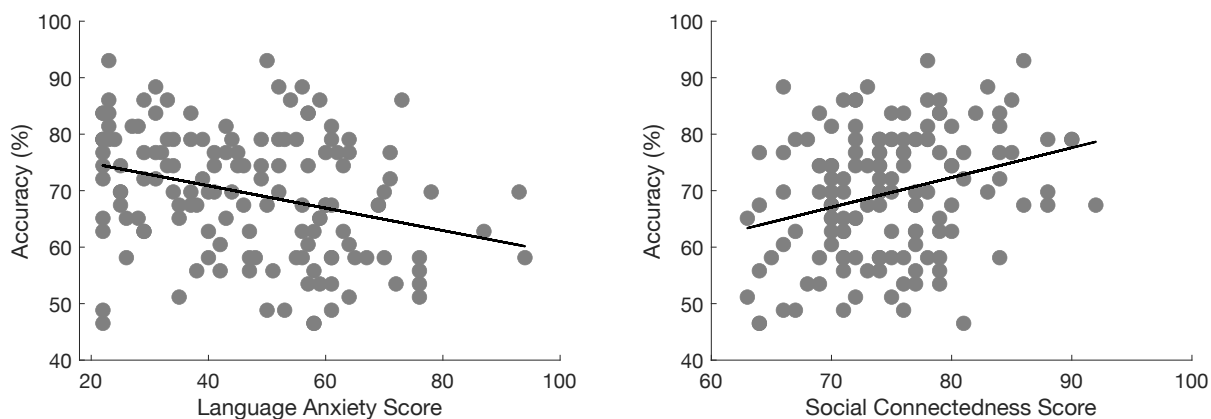
Figure 1: Scatter plots showing the relationship between the accuracy of humor detection and the subjective social well-being scores: (Left) foreign language anxiety score and (Right) the social connectedness score.

to classify each sentence as either a joke or a non-funny statement. After they had completed the humor detection task, participants were asked to complete the three surveys described below: the English Usage and Proficiency Scale, the Foreign Language Anxiety Scale, and the Social Connectedness Scale.

English Usage and Proficiency Scale In total, four questions were asked: "Q1. How long have you been in the United States?", "Q2. What is your present level of English fluency?", "Q3. How comfortable are you in communicating in English?", and "Q4. How often do you communicate in English?". The subjects rated Q1 using a 7-level scale (e.g., 1: less than 1 year, 2: 1 year, ..., and 7: more than 5 years) and the other questions using 5-level scale (e.g., 1: not fluent/uncomfortable/only when necessary, and 5 : fluent/comfortable/for everything). Subjective language fluency scores were computed as the sum of the three self-report scores (i.e., Q2, Q3 and Q4). Language fluency scores thus ranged from 3 to 15 and a higher score indicates greater language fluency.

Foreign Language Anxiety Scale The foreign language anxiety scale was developed to quantify the feeling of anxiety specifically associated with L2 materials (Horwitz et al., 1986) and has been used in multiple studies (Öztürk & Gurbuz, 2014; Park & French, 2013). There are 19 items in the scale. The subjects were asked to rate each of them using 5-level Likert scale (e.g., 1: Strongly disagree, 5: Strongly agree) regarding how well each item describes themselves. The anxiety score was obtained by summing up the responses to each item, leading to the range between 19 and 95. A higher score indicates that the respondent experiences greater levels of language anxiety.

Social Connectedness Scale The social connectedness scale was developed to assess a sense of subjective recognition of being supported or connected to a particular society (Lee et al., 2001). The social connectednes scale has 21 items

in total. We modified some of the original items so that each asked about social connectedness specifically in an English speaking society. The subjects were asked to rate each item using a 5-level Likert scale. The social connectedness score was obtained by summing up the responses to each item, leading to a range between 21 and 105. A higher score indicates that the individual feels a stronger social connection to the English speaking community.

Data Analysis

The accuracy of joke detection was calculated as the number of stimuli that a participant answered correctly divided by the total number of stimuli (i.e., 43). Note that accuracy rates were calculated for each individual participant in this experiment unlike Experiment 1 in which accuracy rates were calculated for each stimulus pair. We then computed the correlation between overall accuracy in the joke detection task with the foreign language anxiety score and the social connectedness score.

In addition, given the nested structure of the data (e.g., repeated measurements within subjects), a mixed effects logistic regression model was applied to better capture the fixed effect of foreign language anxiety, language usage, social connectedness, and type of stimulus (joke vs. non-joke) on predicting accurate classification on each trial. The random effect structure comprised one random intercept term for subject ID and another for Stimulus ID. The joke type was set as reference level. The foreign language anxiety score, language usage score, social connectedness score were normalized to the scale of 0 to 1. Note that for the general form of linear mixed effect regression model, significant factors were those for which the 95% confidence interval (CI) for an odds ratio (OR) does not include 1.

Results

The average accuracy of joke detection across subjects was 69.56 ± 11.10. Figure 1 shows the relationships between the accuracy of humor detection and the subjective social well-

being scores. The accuracy of humor detection was negatively correlated with language anxiety scores ($r = -0.31$, $p < 0.001$), and was positively correlated with social connectedness scores ($r = 0.29$, $p < 0.001$). In addition, language anxiety scores were negatively correlated with social connectedness scores at a moderate level ($r = -0.33$, $p < 0.001$).

The mixed effect logistic regression model revealed a significant effect of non-joke version of the stimulus (OR = 0.09, 95% CI = [0.017, 0.46]), reflecting the fact that participants were more likely to categorize the joke stimuli as jokes than they were to categorize the control stimuli as straightforward statements. Many participants apparently adopted a bias towards the "Joke" response.

Analysis also revealed a significant effect of Language Usage Score (OR = 2.87, 95% CI = [1.62, 5.09]), reflecting better performance on joke classification among participants with higher Language Usage scores. However, the significant interaction of Language Usage and non-joke version (OR = 0.23, 95% CI = [0.12, 0.43]) suggests those same participants were less likely to correctly classify the non-joke versions of stimuli. Taken together, this suggests that the preference to respond "Joke" was greater in participants with higher Language Usage scores.

By contrast, the significant effect of Language Anxiety Score (OR = 0.32, 95% CI = [0.18, 0.56]) reflects lower accuracy in joke classification, with greater accuracy for the non-joke version being reflected in a significant interaction of Language Anxiety score and non-joke version (OR = 2.09, 95% CI = [1.11, 3.92]). Participants who experience greater levels of anxiety speaking English thus exhibited a different response bias, preferring to respond "Statement".

Finally, while Social Connectedness was not associated with a greater probability to respond correctly to jokes, the significant interaction of Social Connectedness Score and non-joke version (OR = 62.16, 95% CI = [7.39, 522.66]) indicates a strong association between this factor and responding correctly to the non-funny version of the stimulus. Given that incorrect responses in this paradigm were most often due to "false alarms" to non-joke stimuli (that is, responding "Joke" to the non-funny version of the stimulus), the large effect of Social Connectedness on accuracy for the statements suggests its relationship to participants' ability to appreciate the difference between the jokes and the non-funny control stimuli.

Discussion

The results confirmed that the ability of humor detection in non-native English speakers is associated with the subjective social well-being in the English speaking community (Figure 1). It seems to be reasonable because humor detection requires a skill to realize intention hidden in sentences. L2 speakers who have lower ability to detect humor might have more chance to face troubles understanding others' intention in conversation, resulting in higher language anxiety and lower social connectedness compared with those who have higher humor competence.

Furthermore, the mixed-effect model indicated that NNS

performed relatively poorer in accurately detecting the non-joke version of the stimulus. More specifically, the interaction term indicated that higher language usage score is associated with lower statement detection accuracy, and higher language anxiety score is associated with higher non-joke stimulus detection accuracy. The opposite effect of language usage score and language anxiety score to joke/non-joke detection accuracy, despite the relatively worse performance in non-joke detection task, suggested that more factors could be associated with humor detection ability. Lastly, a higher sense of social connectedness was shown to improve the non-joke stimulus detection performance.

Experiment 3: Humor appreciation

This experiment aims to examine the relationship between L2 language usage frequency, L2 language anxiety, and subjective social well-being ratings in NNS with their ability to appreciate verbal humor in the L2.

Participants

Participants included 127 NNS subjects (83 females, mean age = 20.43 ± 2.37) after the exclusion of several participants who failed to answer the validation questions correctly. Each subject was compensated with academic course credits. All subjects were at least 18 years old and provided informed consent to participate in the experiment.

Data Collection

The survey format and experimental procedure was similar to that in Experiment 2. However, instead of asking them to perform the joke detection task, subjects in Experiment 3 were asked to rate the funniness of a presented stimulus using 5-level Likert scale (e.g. 1: not funny at all and 5: extremely funny). As in Experiment 2, subjects were also asked to complete the English Usage and Proficiency Scale, the Foreign Language Anxiety Scale, and the Social Connectedness Scale.

Data Analysis

The ratings were averaged across stimuli in each category (i.e., joke and non-joke) for each subject. Initial analysis of the data involved separate correlations between average ratings on joke and non-joke stimuli with Language Anxiety and Social Connectedness, respectively.

The rationale for mixed effects modeling in Experiment 3 was similar to that in Experiment 2. Accordingly, a linear mixed effect model was applied to investigate the fixed effects of foreign language anxiety, language usage, social connectedness, and type of stimulus (joke vs. non-joke) on humor appreciation ratings. The random effect structure comprised one random intercept term for subject ID and another for Stimulus ID. The non-joke version of the stimulus was set as the reference level, since 0 is the normative rating for these stimuli. Note that for the general form of linear mixed effect regression model, significant factors were those for which zero did not fall within the 95% CI.
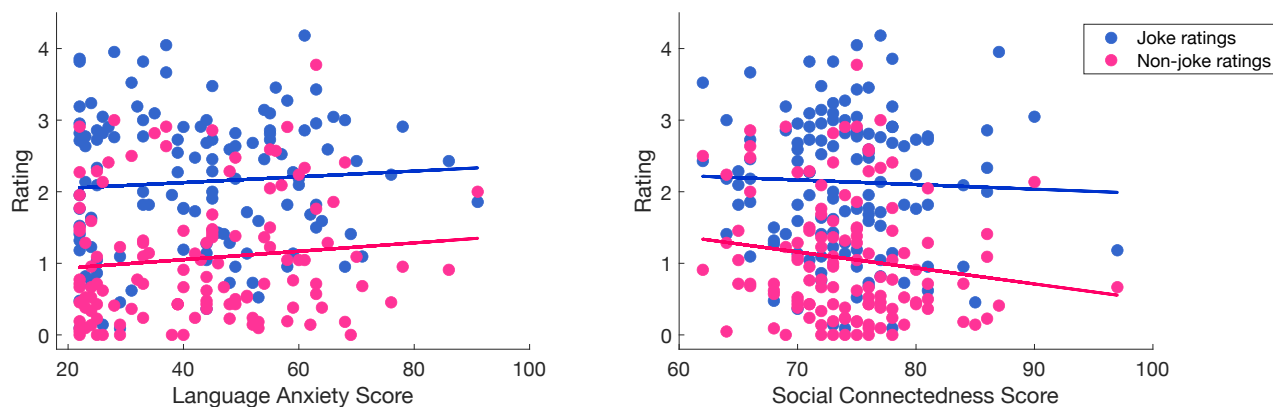
Figure 2: Scatter plots showing the relationship between the humor ratings and the subjective social well-being scores: (Left) foreign language anxiety score and (Right) the social connectedness score. The ratings on joke and non-joke stimuli were shown separately.

Result

Figure 2 shows the relationships between the average ratings across joke and non-joke sentences and the social well-being scores. No significant correlation was found between the ratings and the language anxiety score (joke rating: $r = 0.07$, $p = 0.449$, non-joke rating: $r = 0.11$, $p = 0.222$). Likewise, there were no significant correlations between the ratings and the social connectedness score (joke rating: $r = -0.04$, $p = 0.673$, non-joke rating: $r = -0.15$, $p = 0.100$). As in Experiment 2, there was a significant negative correlation between the language anxiety score and the social connectedness score among participants in Experiment 3 ($r = -0.26$, $p = 0.003$),

The mixed effect regression model revealed significant effect of Language Anxiety (95% CI = [0.37, 1.99]), indicating participants with greater Language Anxiety scores tended to rate the non-funny statement versions of the stimuli as being more humorous than did participants who scored lower on the Language Anxiety scale. Further, an interaction of Joke Version and Language Anxiety Score (95% CI = [-1.17, -0.40]) reflects the fact that participants who scored high on Language Anxiety provided lower humor ratings for the joke stimuli. Finally, the interaction of joke version and Social Connectedness (95% CI = [1.11, 3.72]) reflects the fact that participants with greater Social Connectedness scores assigned higher humor ratings to the joke stimuli.

Discussion

Figure 2 clearly showed different distributions of humor ratings for jokes (median: 2.24, interquartile range: 1.41 - 2.86) and non-jokes (median: 0.77, interquartile range: 0.41 - 1.49), indicating that the subjects tended to rate the jokes as funnier than the non-jokes. The results also revealed that NNS with higher language anxiety score tended to rate non-joke sentences as funnier and joke sentences as less funny, which suggests that NNS with high language anxiety score may experience difficulties on detecting the difference between jokes and non-joke control sentences. In addition, the

positive relationship between social connectedness score and humor rating on joke, confirms our hypothesis that social connectedness is associated with humor appreciation ability.

General Discussion

This study investigated the relationship between L2 humor competence, linguistic competence, and social wellness in the L2 society. Since humor competence requires sophisticated linguistic, social, and cultural competence, we anticipated that the accuracy of humor detection and/or subjective ratings on joke materials of L2 speakers would be associated with their subjective social well-being, their L2 usage frequency, and their sense of anxiety toward L2.

Our results showed that the accuracy of humor detection was significantly associated with the foreign language anxiety score and the social connectedness score. More specifically, language anxiety scores were negatively associated with joke detection accuracy and positively associated with non-joke detection accuracy. One possible reason for the positive association between language anxiety scores and statement detection accuracy is that NNS who have more anxiety about their foreign language competence are more conservative in their judgment that a given statement was intended humorously. By contrast, when unsure regarding the status of a given stimulus, the NNS who were more comfortable with English were more likely to classify it as joke.

It is perhaps worth mentioning that due to the simple design of this study, we could not identify whether the subjects understood the humorous content of these jokes, or whether they based their classifications on the linguistic properties of the sentences. As all of our non-joke stimuli were very similar to jokes, they likely contained many features that are typically diagnostic of verbal humor.

On the other hand, both studies have shown that the Language Anxiety was negatively associated with joke detection and joke appreciation. This consistent finding confirms that NNS with higher language anxiety experience more dif-

ficulty comprehending humorous materials. The consistent positive association between Social Connectedness and joke detection/appreciation across both studies indicates that Social Connectedness is indeed a significant correlate of NNS' ability to accurately detect and react to verbal humor.

Previous empirical investigation of L2 humor comprehension usually relied on subjective rating on humor materials (Ayçiçeği-Dinn, Şişman-Bal, & Caldwell-Harris, 2018; X. Chen & Dewaele, 2019). In such studies, humor detection and appreciation were not explicitly distinguished or humor detection was completely ignored. The present study introduced a novel task that a participant made a judgement whether a given sentence contained any humorous content or not, which provided the accuracy of humor detection. The results indeed showed that humor detection and appreciation were differently associated with the social well-being scales, i.e. the social well-being scores were significantly associated with the accuracy of humor detection but not with the humor ratings. The results confirmed that this approach would be useful in investigating different aspects of humor comprehension. Moreover, the humor detection task might be better to assess the sociolinguistic aspect of humor comprehension than the humor appreciation task.

One limitation of the present study is the lack of an objective measurement of linguistic proficiency. Since it has been known that L2 speakers have a tendency to underestimate their linguistic proficiency in self-assessment (Trofimvich, Isaacs, Kennedy, Saito, & Crowther, 2016), an objective scale would likely provide more reliable assessment of participants' L2 proficiency than the subjective measures employed here. For future studies, one improvement would be revising the question design to dissociate the humor detection and appreciation performance. In the current study, the rating performance could be influenced by the failure of detecting humor in the sentence, hence inducing confound to the result. Nevertheless, the present results clearly support the hypothesis that L2 humor competence is built upon mature linguistic and sociocultural foundations and therefore associated with each other.

References

Ayçiçeği-Dinn, A., Şişman-Bal, S., & Caldwell-Harris, C. L. (2018). Are jokes funnier in one's native language? *Humor*, *31*(1), 5–37.

Bell, N. (2007). Humor comprehension: Lessons learned from cross-cultural communication. *Humor*, *20*(4), 367–387.

Bell, N. D. (2005). Exploring L2 language play as an aid to SLL: A case study of humour in NS–NNS interaction. *Applied Linguistics*, *26*(2), 192-218.

Bell, N. D. (2009). Learning about and through humor in the second language classroom. *Language Teaching Research*, *13*(3), 241–258.

Chen, A., Lin, G.-H., & Yan, H.-W. Y. (2021). Staying connected: Effects of social connectedness, cultural intelli-gence, and socioeconomic status on overseas students' life satisfaction. *International Journal of Intercultural Relations*, *83*, 151-162.

Chen, X., & Dewaele, J.-M. (2019). The relationship between english proficiency and humour appreciation among english L1 users and chinese L2 users of english. *Applied Linguistics Review*, *10*(4), 653–676.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125-132.

Lee, R., Draper, M., & Lee, S. (2001). Social connectedness, dysfunctional interpersonal behaviors, and psychological distress: Testing a mediator model. *Journal of Counseling Psychology*, *48*(3), 310–318.

Nelms, J. L. (2001). *A descriptive analysis of the uses and functions of sarcasm in the classroom discourse of higher education*. University of Florida. ProQuest Dissertations Publishing.

Park, G.-P., & French, B. F. (2013). Gender differences in the foreign language classroom anxiety scale. *System*, *41*(2), 462-471.

Ritchie, G. (1999). Developing the incongruity-resolution theory. In *AISB symposium on creative language: Stories and humor*. Edinburgh.

Shultz, T. R. (1972). The role of incongruity and resolution in children's appreciation of cartoon humor. *Journal of Experimental Child Psychology*, *13*(3), 456–477.

Tarone, E. (2000). Getting serious about language play: Language play, interlanguage variation and second language acquisition. In B. Swierzbin, F. Morris, M. Anderson, C. Klee, & E. Tarone (Eds.), *Social and Cognitive Factors in SLA: Selected Proceedings of the 1999 Second Language Research Forum* (p. 31–54).

Trofimvich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, *19*(1), 122–140.

Öztürk, G., & Gurbuz, N. (2014). Speaking anxiety among turkish efl learners: The case at a state university. *Journal of Language and Linguistic Studies*, *10*(1), 1-17.