**Appendix 1: Description of Coding Categories**

**Learner Profiles**

***Learning context***

This variable was coded as foreign language (FL) or second language (SL). Studies that were coded FL involved participants in countries where the target language was not spoken, whereas studies that were coded as SL involved those in countries where the target language was spoken outside the experiments. Eighty studies were coded as FL, 16 studies were coded as SL, and 3 were coded as mixed (SL+FL). The moderator analysis focused on studies coded as either FL or SL. There was no missing data for this variable.

***L2 proficiency***

This variable was divided into six categories: novice, beginner, lower-intermediate, intermediate, upper-intermediate, and advanced. Novice was defined as participants who had not studied the target language before taking part in the experiment (e.g., Hirata & Kelly, 2010; Silpachai, 2020). Because the proficiency levels were determined in various ways across studies, we set a standard procedure for coding this variable. Decisions on participants’ proficiency levels were based on standardized proficiency test scores reported in studies and/or other objective measures, and when such information was not reported, we relied on researchers’ or instructors’ description of participants’ proficiency. The intercoder agreement for this variable was 98.02%, indicating that we established the reliability of coding this variable. The breakdown of studies coded for L2 proficiency was 16 studies with novice speakers, 14 studies with beginners, 12 studies with lower-intermediate speakers, 8 studies with intermediate speakers, four studies with upper-intermediate speakers, 14 studies with advanced speakers, and 8 studies with mixed proficiency levels. No relevant information was provided for 23 studies (23.2%). The moderator analysis focused on six proficiency levels (novice, beginner, lower-intermediate, intermediate, upper-intermediate, and advanced).

***Age of learning (AOL)***

This variable was defined as the age at which participants studied the target language. The mean and median (if the mean was not reported) of AOL were used for this variable. When the information was not available but the range in AOL was reported, the median was calculated (e.g., when the AOL ranges from 2 to 8, the median is 5). A total of 36 studies reported this information, 52 studies did not (52.5% missing data), and the remaining studies targeted novice learners; hence, 36 data points were analyzed for this variable.

***Age of testing (AOT)***

This variable was defined as the age at which participants took part in the experiment. The mean and median (if the mean was not reported) of AOT were used for this variable. Like AOL, when the information was not available but the range in AOT was provided, the median was calculated from the range values. A total of 91 studies reported this information (8.1% missing data).

**Stimuli Features**

***Target phone***

This variable was divided into five categories: vowel, obstruent, sonorant, syllable structure, and tone. The breakdown of studies for this variable was 50 studies for vowel, 18 studies for obstruent, 11 studies for sonorant, 3 studies for syllable structure, 9 studies for tone, 7 studies for obstruent + sonorant, and 4 studies for vowel + consonant (both obstruent and sonorant). The moderator analysis focused on the five categories (vowel, obstruent, sonorant, syllable structure, and tone). There was no missing data for this variable.

***Inclusion of nonwords***

This variable referred to whether a set of training stimuli contained nonwords (i.e., phonologically plausible words that do not exist in the target language) or not. We coded this variable dichotomously either yes (i.e., training stimuli contained nonwords) or no (i.e., real words). Forty studies used nonwords, 57 studies used real words, and the relevant information was missing in two studies (2%).

**Training Features**

***Task type***

This variable was coded as identification or discrimination tasks. The majority of studies used a forced-choice identification task with various response labels (e.g., keywords, pictures, phonetic symbols). The discrimination tasks consisted of either a same-or-different discrimination task (i.e., AX category discrimination) or a category discrimination task (i.e., AXB or oddity discrimination task). A few studies adopted a mixture of different perception tasks such as AX and AXB discrimination + correct or incorrect identification (Lacabex & Gallardo del Puerto, 2014) or AX and AXB discrimination + yes or no identification (Qian, 2018). Eighty-three studies used identification, seven studies used discrimination, and nine studies used mixed task types. The moderator analysis focused on identification and discrimination tasks. There was no missing data for this variable.

***Response label***

This variable concerned identification training tasks and was coded into four categories: keyword, orthography, phonetic symbol, and visual image. Studies coded as keyword presented word prompts (e.g., *read* vs. *lead*), studies coded as orthography presented fractions of words key to identification (e.g., *r* vs. *l*), studies coded as phonetic symbol presented phonetic notation of the target sounds (e.g., /æ/ vs. /ʌ/), and studies coded as visual image presented non-orthographic symbols or pictures (e.g., flag images, Thomson, 2012). Other types of response labels included Pinyin tonal diacritics (Silpachai, 2020), numbers (Wang, 2013), alphabetic letters (Lambacher et al., 2005), and short versus long (Hirata & Kelly, 2010). Out of 83 studies using identification training tasks, 32 studies used keywords, 19 studies used phonetic symbols, 10 studies used orthography, 5 studies used visual images, 1 study used keywords + phonetic symbols, 17 studies used other types of response labels, and 1 study did not reveal the relevant information (1%). The moderator analysis focused on the four categories (keyword, orthography, phonetic symbol, and visual image).

***Corrective feedback type***

This variable initially consisted of four categories; target, nontarget, combined, and wrong. Studies that provided the correct (i.e., target) forms of sounds as positive feedback when listeners’ responses were incorrect were coded as target. Studies that provided the nontarget forms of sounds as negative feedback were coded as nontarget. Studies that provided both target and nontarget forms as corrective feedback were coded as combined. Studies that only signaled whether listeners’ responses were wrong without exposing them to either target or nontarget forms were coded as wrong. Because there was only one study coded as nontarget (Lee & Lyster, 2016), this category was not included in the analysis. Forty-eight studies provided target feedback, 20 studies provided combined feedback, 24 studies provided wrong-signal feedback, one study provided a mixture of two feedback types in different training tasks (i.e., target for identification, combined for discrimination tasks, Shinohara & Iverson, 2021), and five studies did not provide relevant information (5.1%). The moderator analysis focused on three categories: target, combined, and wrong.

***Talker presentation***

This variable concerned studies that used identification training tasks (k = 83) and consisted of two categories: intermixed and blocked presentation. When participants were presented with a single talker’s voice in an individual training session, such studies were coded as blocked presentation. When participants were presented with two or more talkers’ voices in an individual session, such studies were coded as intermixed presentation. Trial-by-trial talker presentation was blocked in 41 studies and intermixed in 35 studies. Three studies adopted both types of talker presentation and five studies did not provide relevant information (5.1%). The moderator analysis focused on the two categories of blocked and intermixed presentation.

***Environment***

This variable was coded as laboratory, participant-controlled, or classroom environment. When the training schedule was predetermined and participants did not have control of when or where to complete each training session, such studies were labelled as laboratory because training was largely controlled. When participants were allowed some flexibility to choose when or where they completed training sessions (e.g., online, Huensch & Tremblay, 2015), such studies were coded as the participant-controlled condition. When training sessions were part of regular classroom activities or curriculum (e.g., Lacabex & Gallardo del Puerto, 2014), such studies were placed in the classroom environments category. Seventy-six studies were carried out in the laboratory condition, 15 studies were conducted in the participant-controlled condition, seven studies were conducted in the classroom condition, and one study did not report relevant information (1%). The moderator analysis focused on the three categories: laboratory, participant-controlled, and classroom.

***Adaptive training***

This variable was coded dichotomously as either adaptive or fixed training. Fixed training refers to a regular HVPT program in which all participants completed the same number of trials and the same content, whereas in the adaptive mode, the structure of training programs was changed in response to individual learner performance (e.g., Yang et al., 2021). Six studies adopted adaptive training, and 93 studies utilized regular and fixed training. There was no missing data for this variable.

***Number of talkers, target phones, phonetic contexts, and response choices***

Number of talkers was a simple count of native-speaking talkers producing training stimuli. All 99 studies reported this information. Number of target phones was a simple count of L2 sounds trained in the study. For example, when studies aimed to improve identification of /r/ and /l/ contrast (e.g., Lively et al., 1993), the number of phones was two. Number of phonetic contexts was defined as the number of unique words (or syllables) encountered by listeners. For example, when 19 minimal pairs for /i/ and /ɪ/ contrast were used as training stimuli (Giannakopoulou et al., 2013), the number of unique words that listeners encountered during training was 19, and was therefore coded as 19 for this study. Ninety-four studies reported this information (5.1% missing data). Lastly, number of response choices was coded for identification training tasks; however, this was not necessary for discrimination tasks (i.e., AX and ABX category discrimination) because the number of choices was limited to two or three. This variable looked at the number of response choices presented in each trial for listeners to choose from. For example, when *read* and *lead* were presented as response choices on a computer screen, the number of response choices would be two. All 83 studies using identification training tasks reported this information.

**Testing Features**

***Test type***

This variable was coded as either identification or discrimination tests (see also Task type above). Eighty-five studies used identification tests, 16 studies used discrimination tests, and three studies used mixed test types. The moderator analysis focused on the two categories: identification and discrimination tests. There was no missing data for this variable.

***Test item and talker type***

This variable consisted of four categories: old item-old talker, old item-new talker, new item-old talker, and new item-new talker. Old items and talkers mean items or talkers that appeared during both training and pretesting. New items and talkers mean items or talkers that appeared during pretesting, but not during training. Twenty-seven studies used old item-old talker stimuli, 30 studies used old item-new talker stimuli, 18 studies used new item-old talker stimuli, 52 studies used new item-new talker stimuli, and 19 studies did not provide sufficient information to specify the status of test stimuli.

**Modifications to HVPT**

***Audiovisual input***

This variable was coded as providing audiovisual input or audio-only input. In studies providing audiovisual input, participants were presented with a talker’s face (Hardison, 2003) or hand gestures (Hirata & Kelly, 2010) while listening to auditory stimuli. Seven studies provided audiovisual input, and 92 studies provided audio-only input. There was no missing data for this variable.

***Acoustic cue manipulation***

This variable was coded as whether certain acoustic features of the training stimuli were synthetically manipulated or not. For example, with the goal of drawing learners’ attention to vowel quality rather than vowel duration, Wang and Munro (2004) created synthetic stimuli that varied in duration with a view to informing learners that durational cues would not be reliable to distinguish vowel contrasts. Three studies implemented acoustic cue manipulation, and 96 studies did not. There was no missing data for this variable.

**Training Intensity**

***Duration***

This variable was categorically coded into (a) > 2 months, (b) 1 to 2 months, (c) 1 week to 1 month, and (d) < 1 week. Training lasted for more than 2 months in four studies, 1 to 2 months in 28 studies, 1 week to 1 month in 47 studies, less than 1 week in six studies, and the information was not available in 14 studies (14.1%).

***Number of sessions, number of trials per session, and total number of trials***

Number of sessions refers to the number of sets of training trials that participants completed in one go. Ninety-five studies reported this information. Number of trials per session was the number of training trials participants completed in each session. Seventy-five studies reported this information. Total number of trials was the number of training trials participants completed in each experiment. When this information was not available, it was calculated by multiplying the number of trials per session by the number of sessions. Eighty-one studies provided this information (18.1% missing data).

***Training time per session and total training time***

Training time per session was the time participants spent completing training trials in each session (min). Eighty-one studies reported this information. Total training time was the time participants spent completing all training trials in each experiment (min). When this information was not available, it was calculated by multiplying training time per session by the number of sessions. Eighty-two studies provided this information (17.2% missing data).

**References**

Giannakopoulou, A., Uther, M., & Ylinen, S. (2013). Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English. *Child Language Teaching and Therapy*, *29*(2), 201–218. https://doi.org/10.1177/0265659012467473

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522. https://doi.org/10.1017/S0142716403000250

Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*(2), 298–310. https://doi.org/10.1044/1092-4388(2009/08-0243)

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics*, *52*, 105–120. https://doi.org/10.1016/j.wocn.2015.06.007

Lacabex, E. G., & Gallardo del Puerto, F. (2014). Two phonetic-training procedures for young learners: Investigating instructional effects on perceptual awareness. *The Canadian Modern Language Review*, *70*(4), 500–531. https://doi.org/10.3138/cmlr.2324

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227–247. https://doi.org/10.1017/S0142716405050150

Lee, A. H., & Lyster, R. (2016). Effects of different types of corrective feedback on receptive skills in a second language: A speech perception training study. *Language Learning*, *66*(4), 809–833. https://doi.org/10.1111/lang.12167

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255. https://doi.org/10.1121/1.408177

Qian, M. (2018). *An adaptive computational system for automated, learner-customized segmental perception training in words and sentences: Design, implementation, assessment* [Unpublished doctoral dissertation]. Iowa State University.

Shinohara, Y., & Iverson, P. (2021). The effect of age on English /r/-/l/ perceptual training outcomes for Japanese speakers. *Journal of Phonetics*, *89*, 101108. https://doi.org/10.1016/j.wocn.2021.101108

Silpachai, A. (2020). The role of talker variability in the perceptual learning of Mandarin tones by American English listeners. *Journal of Second Language Pronunciation*, *6*(2), 209–235. https://doi.org/10.1075/jslp.19010.sil

Thomson, R. I. (2012). Improving L2 listeners’ perception of English vowels: A computer-mediated approach. *Language Learning*, *62*(4), 1231–1258. https://doi.org/10.1111/j.1467-9922.2012.00724.x

Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal*, *97*(1), 144–160. https://doi.org/10.1111/j.1540-4781.2013.01386.x

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539–552. https://doi.org/10.1016/j.system.2004.09.011

Yang, R., Nanjo, H., & Dantsuji, M. (2021). Self adaptive phonetic training for Mandarin nasal codas. *Computer-Assisted Language Learning Electronic Journal*, *22*(1), 391–413.

**Appendix 2: Equations for Calculating Effect Sizes and Sampling Variance**

**Equation 1**

This equation was used to calculate the standard deviation of the gain scores (*Sdiff*) from the standard deviations of pretest and posttest scores and their correlation.

**Equation 2**

This equation was used to calculate the standard deviation for independent groups (*Swithin*) with the standard deviation for the gain scores (*Sdiff*) and the pretest-posttest correlation (*r* = 0.6).

**Equation 3**

This equation was used to calculate the standardized mean difference between pretest and posttest scores adjusted for the pretest-posttest correlation (*ds*). This notation is used in contrast to Cohen’s *dz*, which is calculated by dividing the gain score (*Mdiff*) by the standard deviation of the gain score (*Sdiff*).

**Equation 4**

**Equation 5**

**Equation 6**

Cohen’s *ds* for gain score was converted to Hedges’ *g* using a correction factor, called *J* with the degrees of freedom (*df* = *n* – 1) and its sampling variance.

**Equation 7**

This equation was used to calculate the pooled standard deviation (*Spooled*) for multiple independent groups.

**Equation 8**

The group mean difference of gain scores (*Mdiff*) was standardized using *Spooled* to calculate Cohen’s *ds*.

**Equation 9**

Cohen’s *ds* for the treatment-control contrast was converted to Hedges’ *g* using a correction factor (*J*) with the degrees of freedom (*df* = *n1* + *n2* – 2) and its sampling variance.

**Equation 10**

**References**

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.

Morris, S. B., & DeShon, R. P. (1997). Correcting effect sizes computed from factor analysis of variance for use in meta-analysis. *Psychological Methods*, *2*(2), 192–199. https://doi.org/10.1037/1082-989X.2.2.192

**Appendix 3: Results of Sensitivity Analysis**

Regarding the pretest-posttest comparison data (k = 99), we performed sensitivity analyses to determine whether the summary effect size and variance estimates substantially vary depending on the different values for within-study correlations used (r = .25, .50, and .75). As summarized in Table 3.1, no substantial changes were observed for the summary effect size (g = 0.915 to 0.917) and variance (tau squared = 0.111 to 0.113). For the treatment-control comparison data (k = 35), Table 3.2 shows no substantial changes in the summary effect size (g = 0.668 to 0.678) and variance (tau squared = 0.010 to 0.012) as a result of using different within-study correlations (r = .25, .50, and .75). Therefore, for both pretest-posttest and treatment-control comparisons, we used r = .50 for computing the summary effect size and its sampling variance.

Table 3.1 Results of sensitivity analysis for pretest-posttest comparison

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | K | M(g) | SE | 95% CI | Tau Squared |
| R = .25 | 96 | 0.917 | 0.043 | [0.832, 1.002] | 0.113 |
| R = .50 | 96 | 0.916 | 0.044 | [0.830, 1.002] | 0.112 |
| R = .75 | 96 | 0.915 | 0.045 | [0.827, 1.002] | 0.111 |

Table 3.2 Results of sensitivity analysis for treatment-control comparison

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | K | M(g) | SE | 95% CI | Tau Squared |
| R = .25 | 32 | 0.668 | 0.059 | [0.552, 0.784] | 0.012 |
| R = .50 | 32 | 0.673 | 0.061 | [0.554, 0.792] | 0.011 |
| R = .75 | 32 | 0.678 | 0.062 | [0.556, 0.799] | 0.010 |

**Appendix 4: Inspection of Influential Cases**

Regarding the pretest-posttest comparison data (k = 99), we conducted a leave-one-out analysis. We did not use any cut-off criteria to determine outliers as a primary approach. This traditional approach is not suitable for the current meta-analysis using a random- and mixed-effects model (instead of fixed-effects models), dealing with two sources of variability (i.e., sampling variability and between-study heterogeneity) in the same model (see Viechtbauer & Cheung, 2010). The leave-one-out analysis performs the calculation of the mean effect size multiple times by excluding one study at each analysis; thus, we can investigate the influence of each study on the overall effect-size estimate and identify overly influential studies. Figure 4.1 is a plot of the weighted mean while excluding each of the 99 studies included in the meta-analysis. As presented in this figure, this analysis revealed that two cases had a major impact on the overall mean effect size in the current data set (i.e., g = 6.60 in Y. Li, 2015 and g = 4.21 in Qian, 2018) that may potentially lead to the overestimation of the true effect. Thus, we excluded these two studies from the subsequent analysis. We also excluded one study (i.e., Georgiou, 2021) that showed an extremely high effect size (g = 10.4) almost 10 times larger than the mean effect size (g = 1.02). Although the leave-one-out analysis did not seem to show that it was a clear influential case, this may be because this study was not given as much weight due to its high sampling variance (variance = 2.93). Regarding the treatment-control comparison (k = 35), we also conducted a leave-one-out analysis and confirmed that three studies were clear influential cases (i.e., g = 6.34 in Y. Li, g = 1.93 in Qian, 2018, and g = 3.02 in Wong, 2013), potentially leading to overestimation of the true effect. Therefore, we excluded these three studies for the subsequent analysis.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Figure 4.1

Plot of the weighted means with each of the 99 effect sizes excluded at each analysis (pretest-posttest comparison), indicating that two cases (Y. Li, 2015 and Qian, 2018) led to the overestimation of the true effect

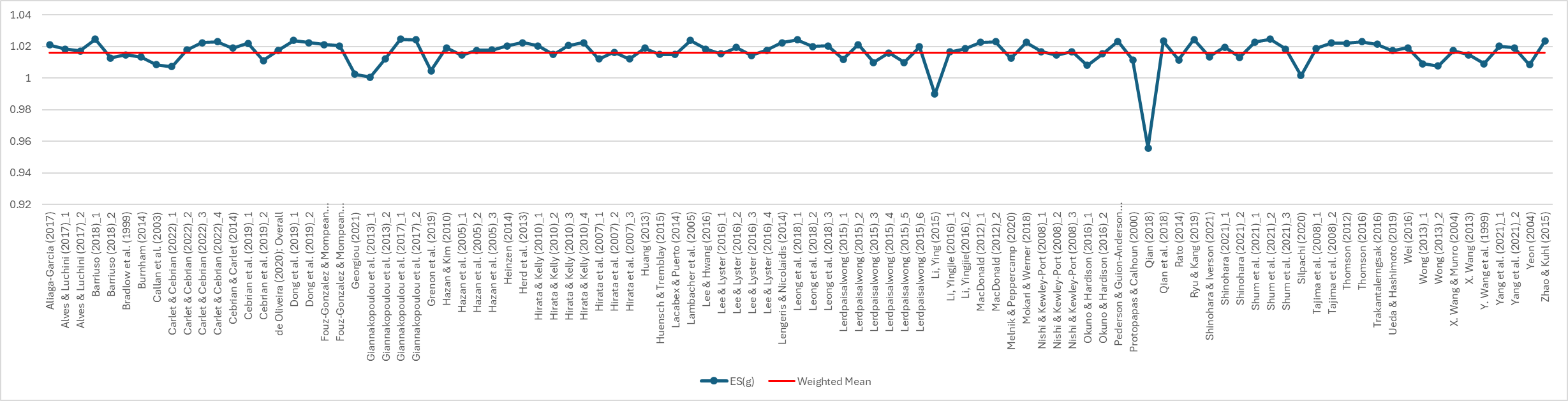
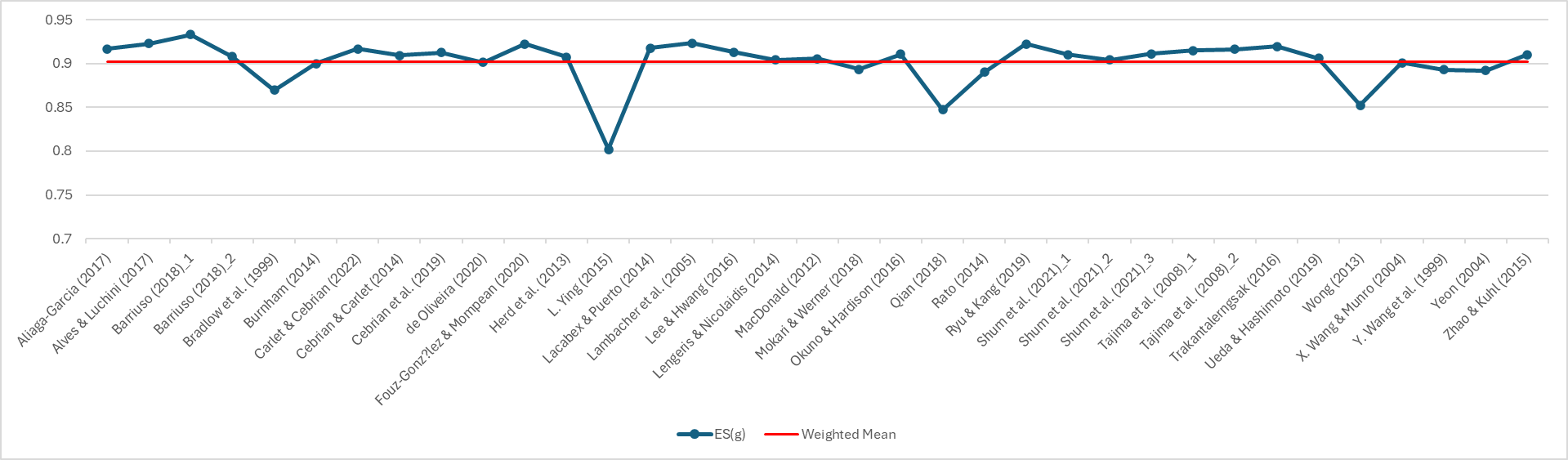


Figure 4.2

Plot of the weighted means with each of the 35 effect sizes excluded at each analysis (pretest-posttest comparison), indicating that three cases (Y. Li, 2015, Qian, 2018, and Wong, 2013) led to the overestimation of the true effect



**Appendix 5: Forest Plots of 96 Studies (Pretest-Posttest Contrast) and 32 Studies (Treatment-Control Contrast) Included in the Analysis**

Figure 5.1

Forest plot of 96 studies (pretest-posttest comparison)

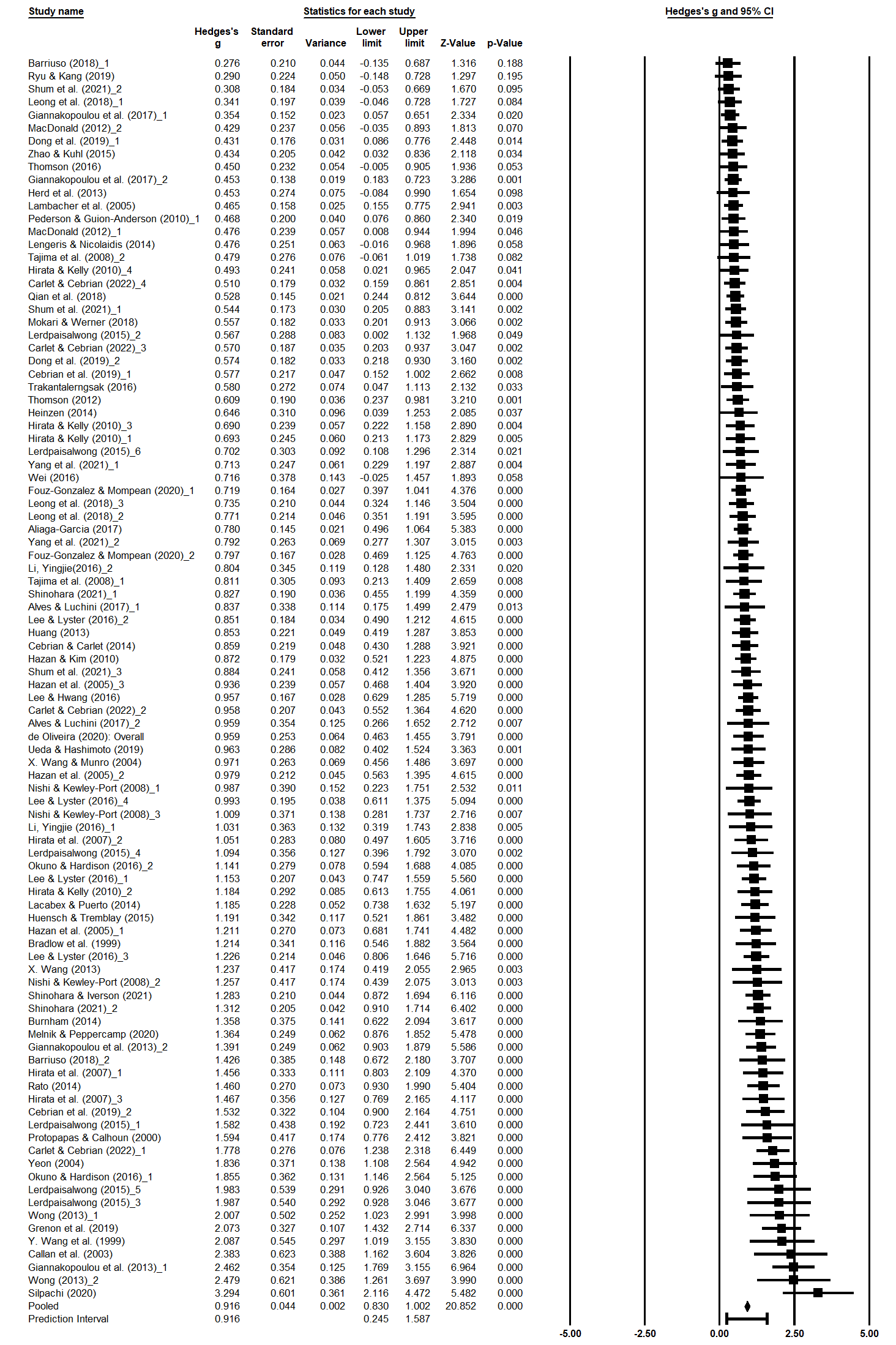
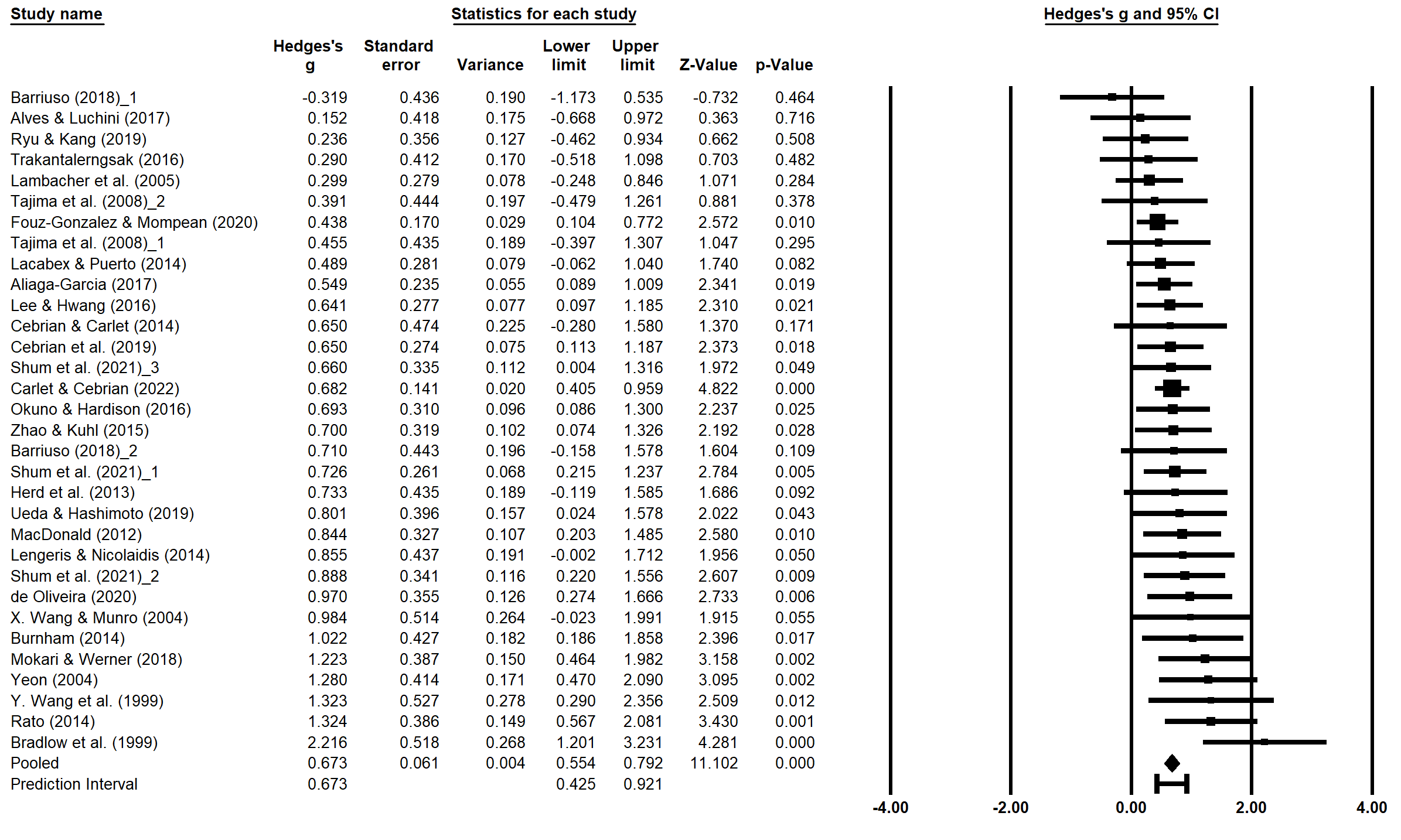
****

Figure 5.2

Forest plot of 32 studies (treatment-control comparison)



**Appendix 6: Detailed Summaries of Results for Retention (RQ1b), Generalization (RQ1c) and Control Group (RQ3)**

**RQ1b: Is the Perception Accuracy Gain Retained?**

In order to determine whether perception accuracy gain from HVPT can be retained over time, we conducted two analyses of effect-size aggregation for the comparison of pretest versus delayed posttest scores and posttest versus delayed posttest scores. A total of 30 studies reported perception accuracy scores at pretest, posttest, and delayed posttest. Thirty studies consist of 18 journal articles, nine PhD theses, and three proceedings reports. The average sample size was 18.8 (range = 8 to 74). The majority of participants at the time of testing were university students (k = 24) and studied target languages in foreign language (FL) contexts (k = 23) in contrast to a few studies focusing on learners in second language (SL) contexts (k = 7). Studies were conducted in various countries and regions such as Argentina (k = 2), Canada (k = 6), China (k = 1), Hong Kong (k = 3), Japan (k = 1), Malaysia (k = 2), Portugal (k = 1), Spain (k = 8), Taiwan (k = 1), the UK (k = 3), and the USA (k = 3). Most studies focused on English as the target language (k = 23) while others focused on French (k = 2) or Mandarin Chinese (k = 2). Out of 21 studies reporting information about the participants’ L2 proficiency, it ranged from novice (k =1), beginner (k = 3), lower-intermediate (k = 1), intermediate (k = 1), upper-intermediate (k = 1), to advanced (k = 9) with the remainder as mixed (k = 4). Regarding target phones, more than half of the studies examined the accuracy at perceiving vowels (k = 15) followed by obstruents (k = 7), tones (k = 2), syllable structures (k = 1), sonorants (k = 1), and other mixed sounds (e.g., vowels + consonants, k = 4). Twenty-three studies adopted identification tasks, five studies used discrimination tasks, and the remaining two studies used mixed tasks. The mean interval between posttests and delayed posttests was 2.3 months (SD = 1.8, range = 0.5 to 6 months). Other basic features of the included studies pertaining to training procedures and intensity (e.g., number of talkers, trials, and sessions) are summarized in Table 6.1.

Table 6.1. Summary of training procedures and intensity

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | k | M | SD | Range |
| Number of talkers | 30 | 4.5 | 1.7 | 2 – 12 |
| Number of training sessions | 29 | 8.8 | 8.4 | 3 – 45 |
| Total number of trials | 24 | 2057.9 | 1196.1 | 792 – 6120 |
| Total training time (min) | 27 | 324.1 | 240.0 | 60 – 1125 |
| Number of trials per session | 22 | 283.0 | 155.7 | 72 – 576 |
| Training time per session (min) | 27 | 38.0 | 20.6 | 5 – 75 |

The first results (i.e., pretest vs. delayed posttest performance) showed that the mean effect size was significant and large, k = 30, g = 0.98, SE = 0.10, 95% CI[0.79, 1.17], p < .001, indicating that delayed posttest scores were larger than pretest scores. The heterogeneity between the studies was significant and high, Q(29) = 105.97, p < .001, I2 = 72.6%, indicating that only about 30% of the heterogeneity was due to sampling error. The 95% prediction interval ranged from 0.078 to 1.886, indicating that with 95% probability a future observation of the training effect on retention (i.e., pretest vs. delayed posttest comparison) will fall in this interval. The result of a leave-one-out analysis did not reveal any clear outliers that had a substantial impact on the overall mean effect size (see Figure 6.1). Publication bias was first visually assessed with a funnel plot (Figure 6.2), indicating a tendency that studies with a lower precision (or higher SE) produced larger effect sizes. To further quantify this potential publication bias asymmetry, we conducted Egger’s test, confirming the presence of publication bias, t(28) = 4.57, p < .001. The trim-and-fill method identified 9 samples that could be hypothetically added to retrieve the symmetry shape of the data distribution, resulting in a smaller estimate of the mean effect size, g = 0.73, 95% CI[0.53, 0.94]. These results indicate the possibility that the observed mean effect was overestimated due to the influence of publication bias. The forest plot for 30 studies is presented in Figure 6.3.

Figure 6.1

Plot of the weighted means with each of the 30 studies excluded at each analysis (pretest vs. delayed posttest comparison)

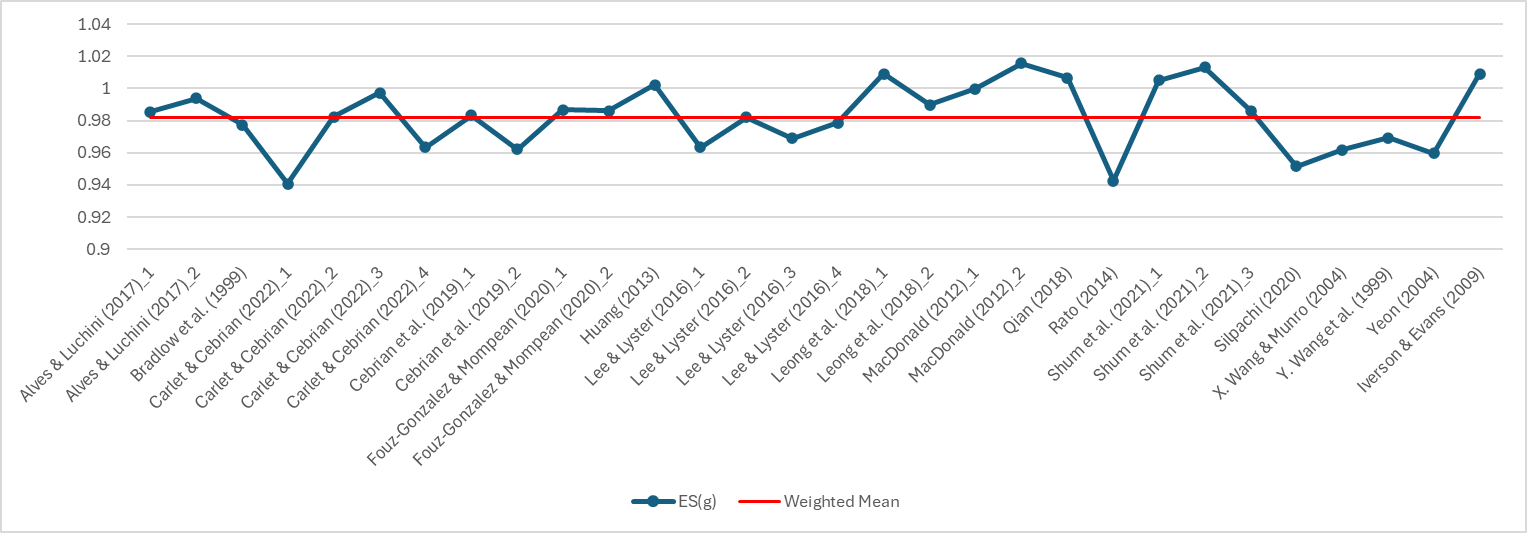


Figure 6.2

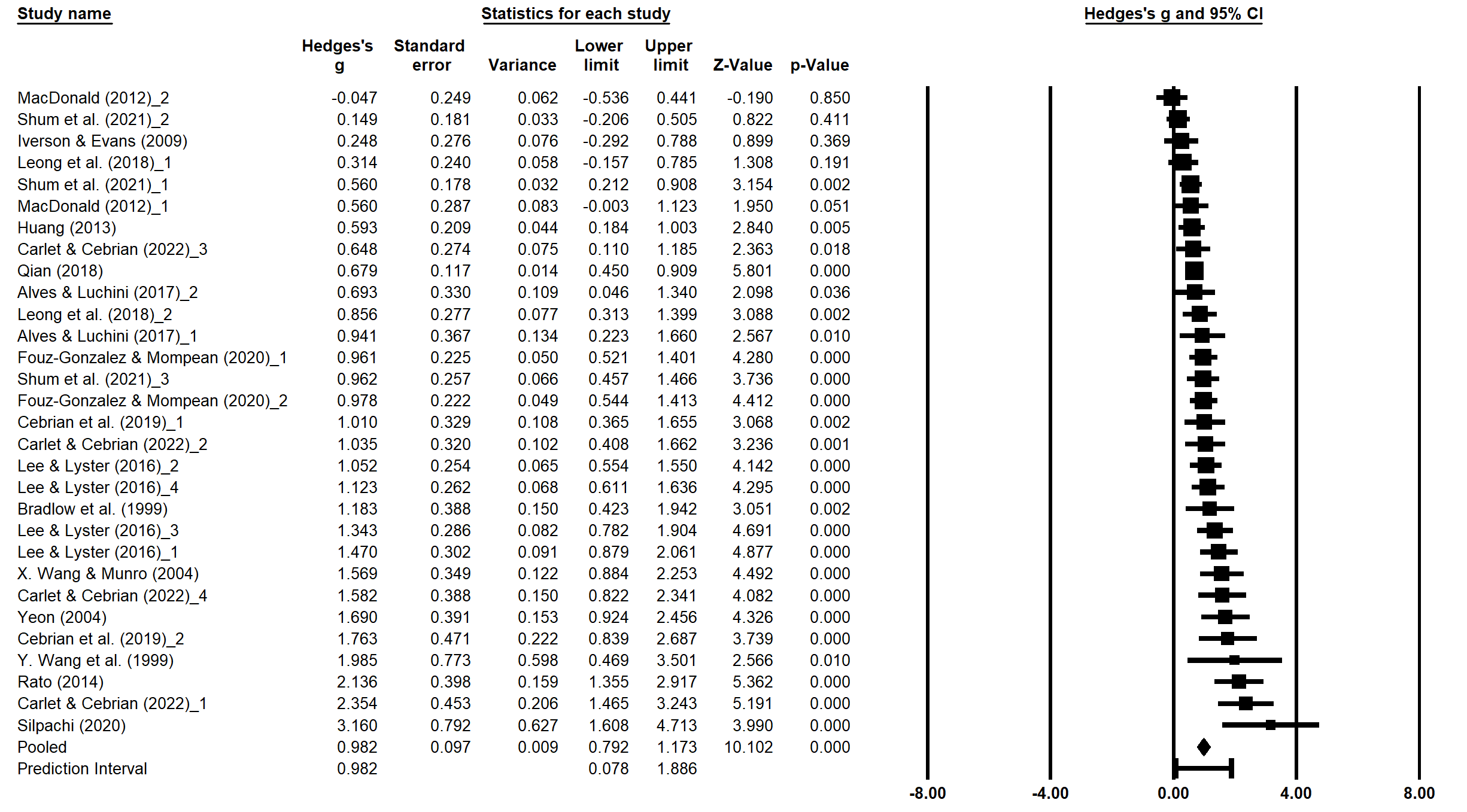
Funnel plot of perception effect size (pretest vs. delayed posttest) by inverse standard error

ダイアグラム

低い精度で自動的に生成された説明

Figure 6.3

Forest plot of 30 studies included in the meta-analysis (pretest vs. delayed posttest)



The second results (i.e., immediate vs. delayed posttest) showed that the mean effect size approached statistical significance but was negligible, g = –0.08, SE = 0.04, 95% CI[–0.16, 0.00], p = .058. The heterogeneity between the studies was nonsignificant and low, Q(29) = 36.33, p = .164, I2 = 20.2%, indicating that only about 20% of the observed variance reflects variance in true effects (i.e., 80% of the observed variance was due to sampling error). The 95% prediction interval ranged from -0.304 to 0.146, indicating that with the 95% probability a future observation of the training effect on retention (i.e., immediate vs. delayed posttest comparison) will fall in this interval. The result of a leave-one-out analysis did not reveal any clear outliers that had a substantial impact on the overall mean effect size (see Figure 6.4). Publication bias was visually assessed with a funnel plot (Figure 6.5). No clear association between precision and effect sizes was observed, which was confirmed with the nonsignificant result of the Egger’s test, t(28) = 0.67, p = .509. The forest plot for 30 studies is presented in Figure 6.6.

Figure 6.4

Plot of the weighted means with each of the 30 studies excluded at each analysis (immediate vs. delayed posttest comparison)

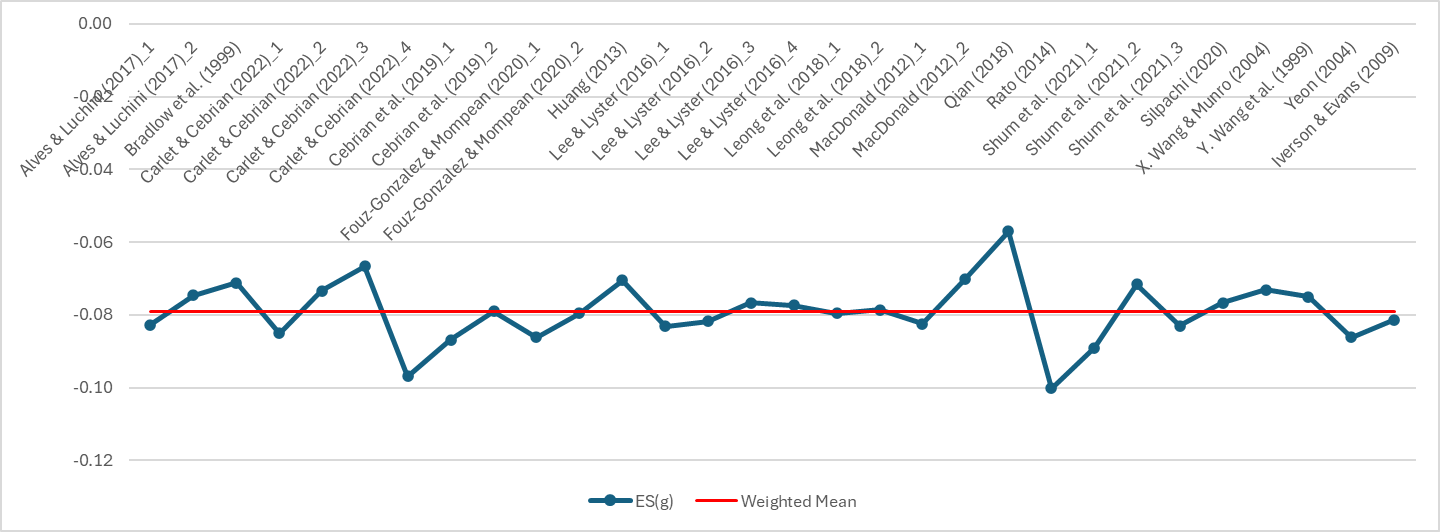


Figure 6.5

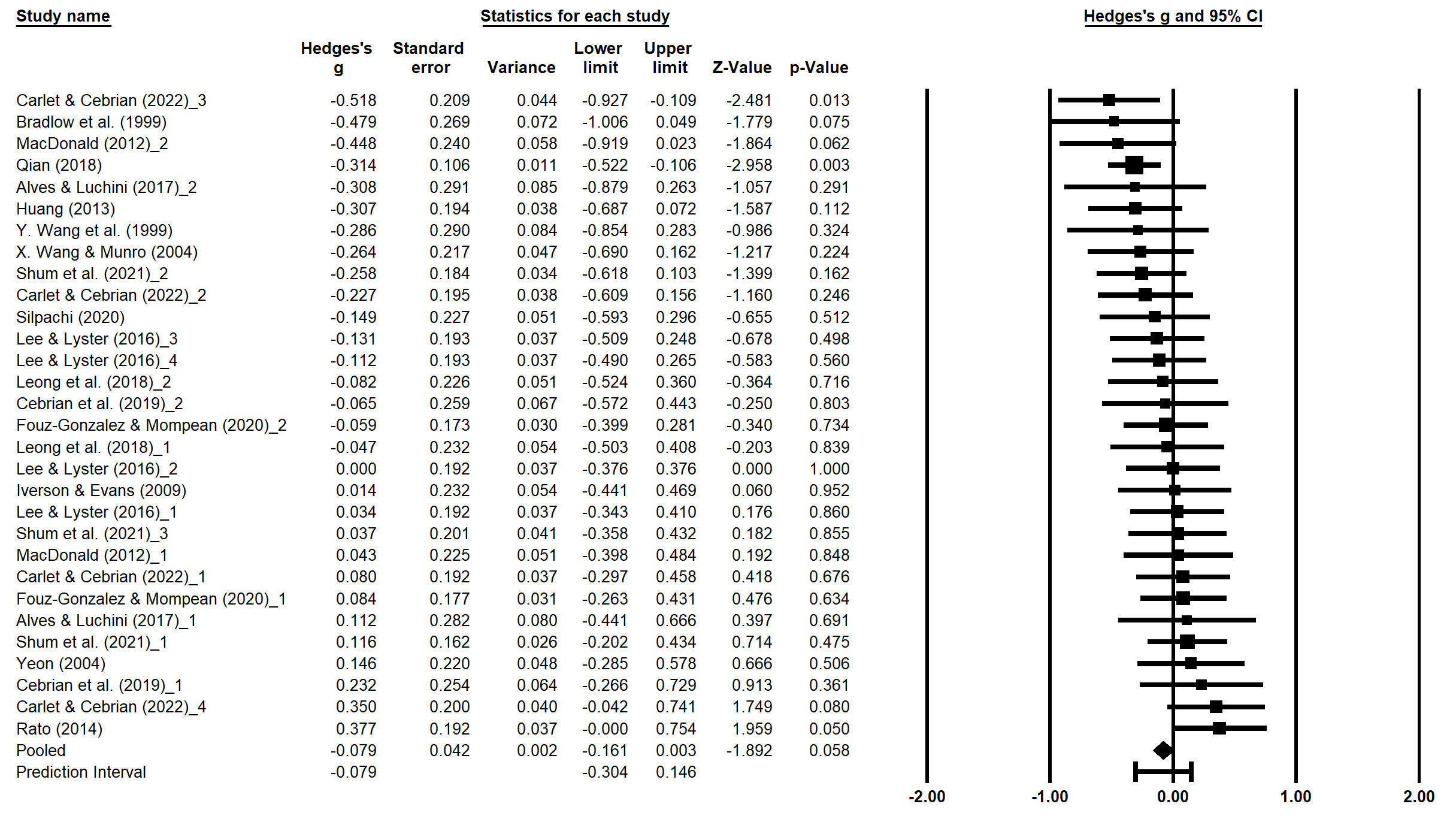
Funnel plot of perception effect size (immediate vs. delayed posttest) by inverse standard error

ダイアグラム

中程度の精度で自動的に生成された説明

Figure 6.6

Forest plot of 30 studies included in the meta-analysis (immediate vs. delayed posttest)



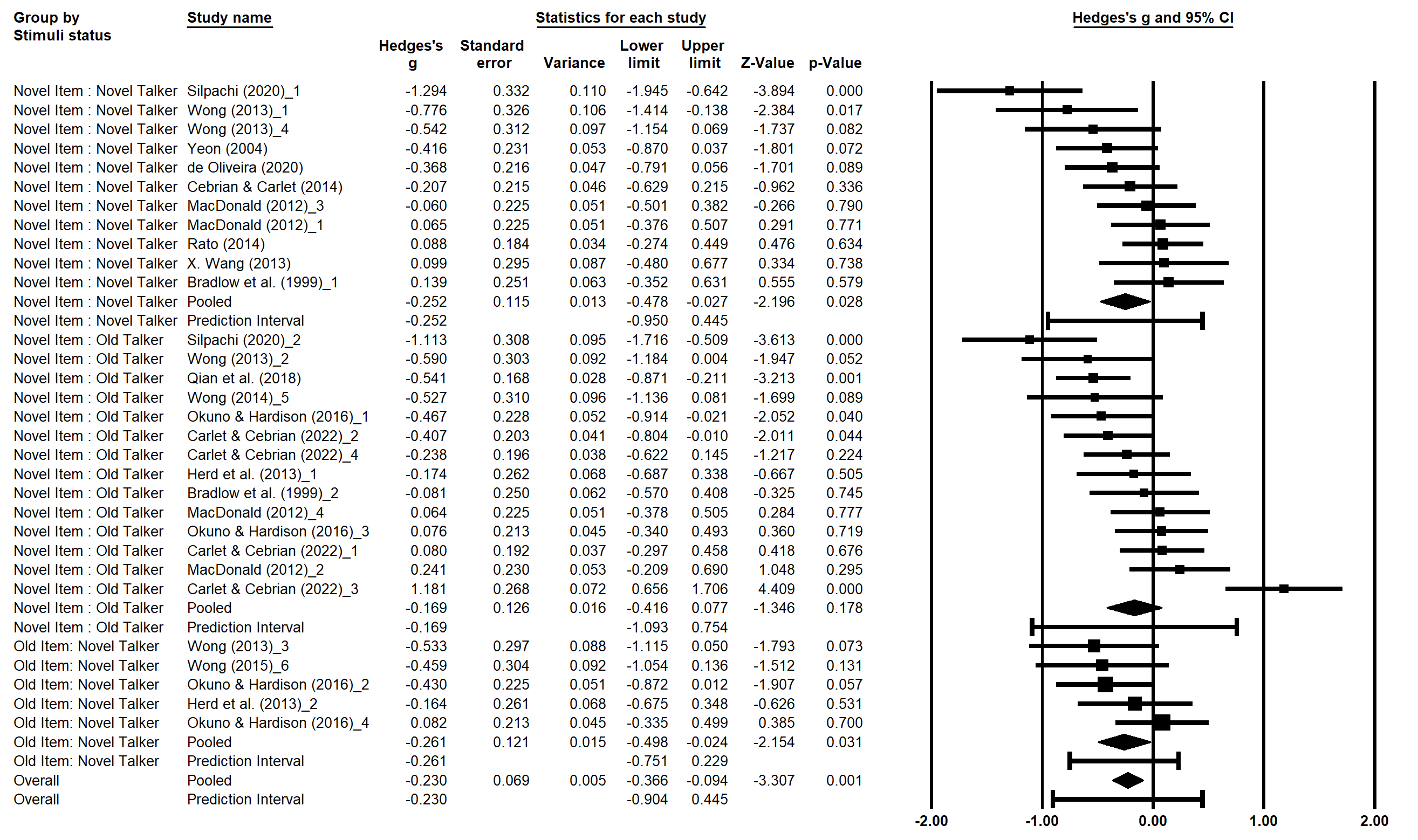
**RQ1c: Is the Improved Perception Generalized?**

We aggregated the effect sizes (i.e., the mean difference between posttest and generalized posttest scores) for each of the three categories of item and talker types: (1) novel item and novel talker (k = 12), (2) novel item and old talker (k = 14), and (3) old item and novel talker (k = 5). Our goal of this analysis was to examine whether there would be any difference between the three categories rather than calculating the average effect of training (for the overall mean effect of HVPT, see the results reported in answer to RQ1a). In general, the majority of included studies came from journal articles (k = 13) and PhD theses (k = 16) with the remainder from proceedings reports (k = 2). Over 90% of the studies focused on participants in FL contexts (k = 29) and experiments were conducted in various countries and regions such as Hong Kong (k = 6), Japan (k = 3), Portugal (k = 2), Russia (k = 1), Spain (k = 5), the UK (k = 4), and the US (k = 10). English is a commonly targeted language (k = 17), followed by Japanese (k = 4), French (k = 4), Mandarin Chinese (k = 3), Spanish (k = 2), and Portugues (k = 1). The average sample size ranged from 11.8 (old item & novel talker), 13.3 (novel item & novel talker), to 16 (novel item & old talker).

The initial result showed no significant variation across three categories, Q(2) = 0.89, p = .642. We identified one study (g = –6.14 in Ueda & Hashimoto, 2019) as a clear outlier that reported the effect size more than six times larger than any of the mean effects calculated in each of the three categories (M = –0.36 to –0.17). After this study was excluded (novel item & novel talker: k = 11), the between-category variability remained nonsignificant, Q(2) = 0.34, p = .845. A further examination of the data showed that the mean effect size was not significant for the category of novel item and old talker, g = –0.17, SE = 0.13, 95% CI[–0.42, 0.08], p = .178. The mean effect size was significant but small for the category of novel item and novel talker, g = –0.25, SE = 0.12, 95% CI[–0.48, –0.03], p = .028 and for the category of old item and novel talker, g = –0.26, SE = 0.12, 95% CI[–0.50, –0.02], p = .031. These results are summarized in the forest plot (see Figure 6.7).

Figure 6.7

Forest plot of 30 studies included in the meta-analysis (comparison of three categories of test item and talker types).



**RQ3: Do Test-Only Control Groups Improve Perception Accuracy?**

We examined the degree to which control groups improved by taking tests without completing training sessions or receiving training with nontarget sounds (k = 31). The included studies appeared in journal articles (k = 17), PhD theses (k = 10), and proceedings reports (k = 4). Most studies were conducted in FL contexts (k = 27); the remaining studies were conducted in SL contexts (k = 4). These experiments were conducted in various countries and regions, including Argentina (k = 1), Canada (k = 4), China (k = 2), Hong Kong (k = 4), Japan (k = 3), Korea (k = 1), Portugal (k = 1), Spain (k = 6), Thailand (k = 1), the UK (k = 1), and the US (k = 7). English was most often examined as a target language, followed by Japanese (k = 4), Arabic (k = 1), Korean (k = 1), Mandarin Chinese (k = 1), and Spanish (k = 1). Out of 26 studies which reported participants’ educational levels, 20 studies focused on university students (k = 20), followed by secondary (k = 3) and primary (k = 3) students. Twenty studies which indicated relevant information about participants’ L2 proficiency focused on learners with various L2 proficiency levels: advanced (k = 5), upper-intermediate (k = 2), intermediate (k = 5), lower-intermediate (k = 1), and beginner (k = 4), with the remainder as mixed (k = 1). Regarding target phones, about half of the studies examined the accuracy at perceiving vowels (k = 14) followed by obstruents (k = 9), tones (k = 1), syllable structures (k = 1), sonorants (k = 2), and other mixed sounds (e.g., vowels + consonants, k = 4). Perception pretests and posttests were administered using identification tasks in most studies (k = 24) and either discrimination tasks (k = 3) or combined tasks (k = 4) in the other studies. Out of 27 studies indicating relevant information, the pretest and posttest interval was approximately 1 week to 1 month in 12 studies, 1 to 2 months in 13 studies, and 2 months or longer in two studies. Twenty-six studies adopted test-only control conditions where participants did not complete any training trials while five studies completed training trials that did not involve target phonological features.

The aggregation of the 31 effect sizes showed that the mean effect size was significant but small, g = 0.19, SE = 0.07, 95% CI[0.05, 0.33], p = .009. The heterogeneity between the studies was significant and relatively high, Q(30) = 86.18, p < .001, I2 = 65.2%, indicating that about 35% of the heterogeneity was due to sampling error. The 95% prediction interval ranged from –0.478 to 0.853, indicating that with the 95% probability a future observation of the practice effect will fall in this interval. The result of a leave-one-out analysis did not reveal any clear outliers that had a substantial impact on the overall mean effect size (see Figure 6.8). Publication bias was first visually assessed with a funnel plot (Figure 6.9). No clear association between precision and effect sizes was observed, which was confirmed with the nonsignificant result of the Egger’s test, t(29) = 0.009, p = .993. The forest plot for 31 studies is presented in Figure 6.10.

Figure 6.8

Plot of the weighted means with each of the 31 studies excluded at each analysis (pretest vs. posttest of control groups)

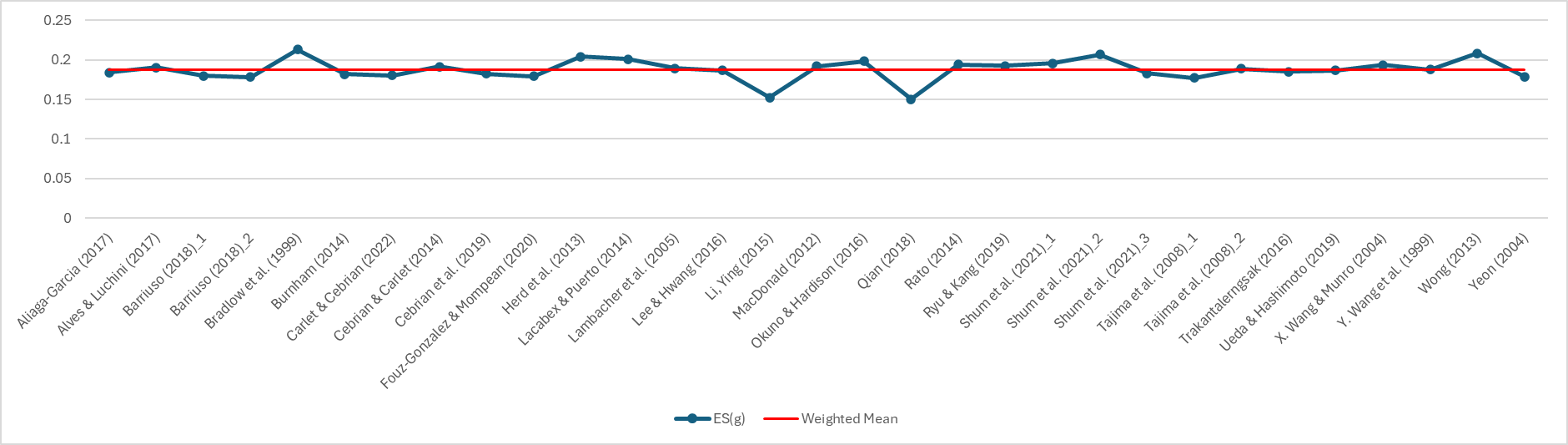


Figure 6.9

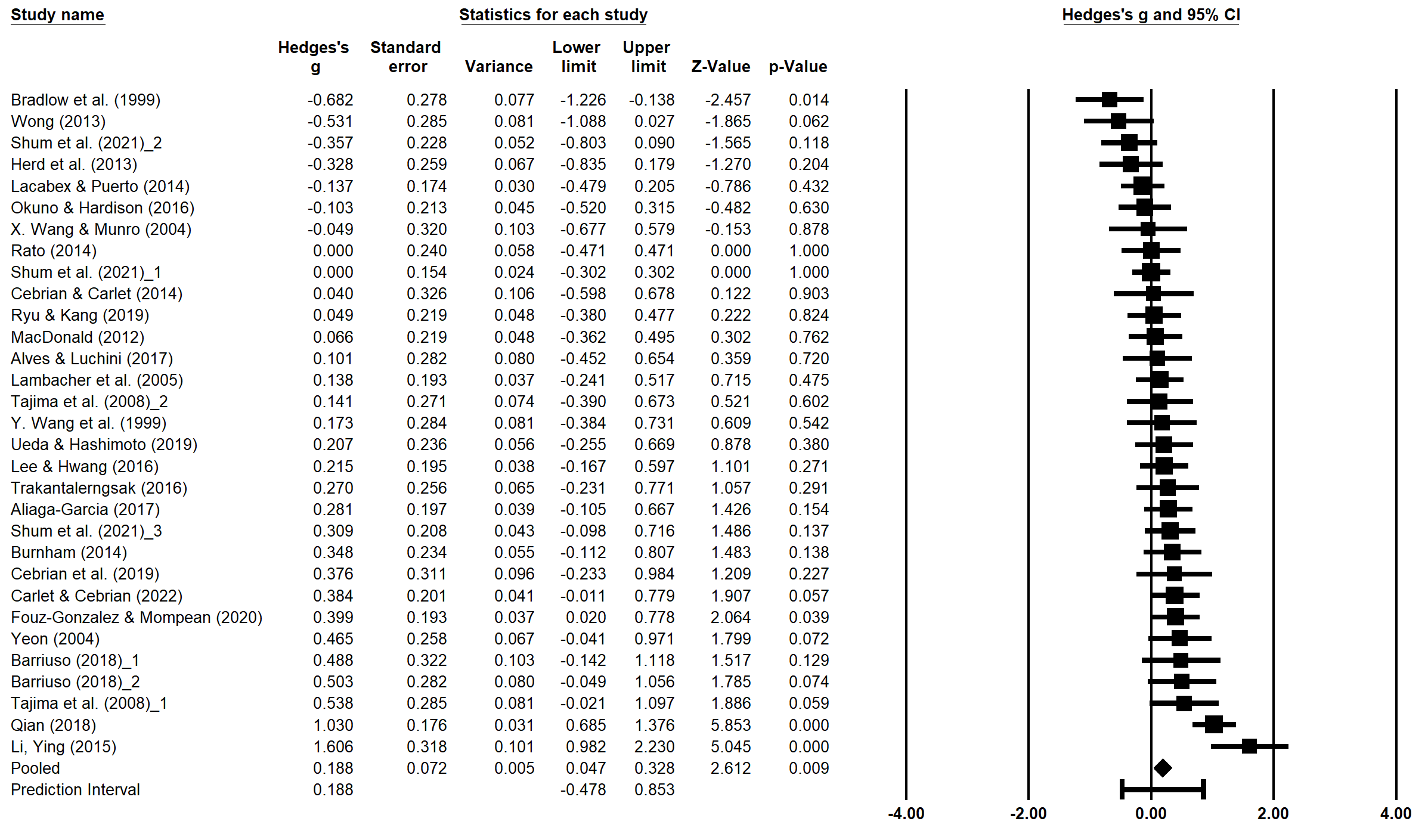
Funnel plot of perception effect size (pretest vs. posttest of control groups) by inverse standard error

ダイアグラム

自動的に生成された説明

Figure 6.10

Forest plot of 31 studies included in the meta-analysis (pretest vs. posttest of control groups)



**Appendix 7: Full References for 79 Studies in Meta-Analysis**

Aliaga-Garcia, C. (2017). *The effect of auditory and articulatory phonetic training on the perception and production of L2 vowels by Catalan-Spanish learners of English* [Unpublished doctoral dissertation]. Universitat de Barcelona.

Alves, U.K., & Luchini, P.L. (2017). Effects of perceptual training on the identification and production of word-initial voiceless stops by Argentinean learners of English. *Ilha Do Desterro*, 70(3), 15 – 32. https://doi.org/10.5007/2175-8026.2017v70n3p15

Barriuso, T. A. (2018). *The L2 acquisition of phonemes and allophones under various exposure conditions* [Unpublished doctoral dissertation]. The University of Utah.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977–985. https://doi.org/10.3758/BF03206911

Burnham, K. R. (2013–2014). Phonetic training in the foreign language curriculum. *Applied Language Learning*, *23–24*, 63–74.

Callan, D.E., Tajima, K., Callan, A.M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage* 19, 113 – 124. <https://doi.org/10.1016/S1053-8119(03)00020-X>

Carlet, A., & Cebrian, J. (2022). The roles of task, segment type, and attention in L2 perceptual training. Applied Psycholinguistics, 43(2), 271–299. <https://doi.org/10.1017/S0142716421000515> & Carlet, A. (2017). *L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross-training study* [Unpublished doctoral dissertation]. Universitat Autònoma de Barcelona.

Cebrian, J., & Carlet, A. (2014). Second-language learners’ identification of target-language phonemes: A short-term phonetic training study. *The Canadian Modern Language Review*, 70(4), 474 – 499. <https://doi.org/10.3138/cmlr.2318>

Cebrian, J., Carlet, A., Gorba, C., & Gavaldà, N. (2019). Perceptual training affects L2 perception but not cross-linguistic similarity. *Proceedings of the 19th International Congress of Phonetic Sciences*, 929 – 933.

de Oliveira, D. M. (2020). *Auditory selective attention and performance in high variability phonetic training: The perception of Portuguese stops by Chinese L2 learners* [Unpublished doctoral dissertation]. Universidade do Minho.

Dong, H., Clayards, M., Brown, H., & Wonnacott, E. (2019). The effects of high versus low talker variability and individual aptitude on phonetic training of Mandarin lexical tones. *PeerJ*, *7*, e7191. https://doi.org/10.7717/peerj.7191

Flege, J. E. (1995). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*(4), 425–442. <https://doi.org/10.1017/S0142716400066029>

Fouz-González, J., & Mompean, J. A. (2021). Exploring the potential of phonetic symbols and keywords as labels for perceptual training. *Studies in Second Language Acquisition*, *43*(2), 297–328. <https://doi.org/10.1017/S0272263120000455>

Francis, A.L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36, 268 – 294. <https://doi.org/10.1016/j.wocn.2007.06.005>

Georgiou, G. P. (2021). Effects of phonetic training on the discrimination of second language sounds by learners with naturalistic access to the second language. *Journal of Psycholinguistic Research, 50*(3), 707–721. https://doi.org/10.1007/s10936-021-09774-3

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, *5*, e3209. <https://doi.org/10.7717/peerj.3209>

Giannakopoulou, A., Uther, M., & Ylinen, S. (2013). Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English. *Child Language Teaching and Therapy*, *29*(2), 201–218. https://doi.org/10.1177/0265659012467473

Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. Journal of Phonetics, 72, 17 – 34. <https://doi.org/10.1016/j.wocn.2018.10.005>

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522. <https://doi.org/10.1017/S0142716403000250>

Hazan, V., & Kim, Y. H. (2010). Can we predict who will benefit from computer-based phonetic training? In *Online Proceedings of the INTERSPEECH 2010 Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology* (L2WS 2010), Tokyo, Japan.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3), 360 – 378. <https://doi.org/10.1016/j.specom.2005.04.007>

Heinzen, C.C. (2014). *Brain plasticity in speech training in native English speakers learning Mandarin Tones* (Unpublished master thesis). University of Minnesota.

Herd, W., Jongman, A., & Sereno, J. (2013). Perceptual and production training of intervocalic /d, ɾ, r/ in American English learners of Spanish. *The Journal of the Acoustical Society of America*, *133*(6), 4247–4255. <https://doi.org/10.1121/1.4802902>

Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *The Journal of the Acoustical Society of America*, 116, 2384 – 2394. <https://doi.org/10.1121/1.1783351>

Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*(2), 298–310. <https://doi.org/10.1044/1092-4388(2009/08-0243)>

Hirata, Y., Whitehurst, E., & Cullings, E. (2007). Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. *The Journal of the Acoustical Society of America*, *121*(6), 3837–3845. <https://doi.org/10.1121/1.2734401>

Huang, Y.F. (2013). *The effects of two methods on training EFL university students in Taiwan to identify three non-native phonetic contrasts* (Unpublished doctoral dissertation). The Ohio State University.

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics*, *52*, 105–120. https://doi.org/10.1016/j.wocn.2015.06.007

Iino, A., & Thomson, R.I. (2018). Effects of web-based HVPT on EFL learners’ recognition and production of L2 sounds. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouesny (Eds.), *Future-proof CALL: Language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 106 – 111). <https://research-publishing.net/book?10.14705/rpnet.2018.26.9782490057221>

Iverson, P., & Evans, B.G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, 126, 866 – 877. <https://doi.org/10.1121/1.3148196>

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278. <https://doi.org/10.1121/1.2062307>

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(1), 145–160. <https://doi.org/10.1017/S0142716411000300>

Lacabex, E. G., & Gallardo del Puerto, F. (2014). Two phonetic-training procedures for young learners: Investigating instructional effects on perceptual awareness. *The Canadian Modern Language Review*, *70*(4), 500–531. <https://doi.org/10.3138/cmlr.2324>

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227–247. https://doi.org/10.1017/S0142716405050150

Lee, H.Y., & Hwang, H. (2016). Gradient of learnability in teaching English pronunciation to Korean learners. *The Journal of the Acoustical Society of America*, 139, 1859 – 1872. <https://doi.org/10.1121/1.4945716>

Lee, A. H., & Lyster, R. (2016). Effects of different types of corrective feedback on receptive skills in a second language: A speech perception training study. *Language Learning*, *66*(4), 809–833. https://doi.org/10.1111/lang.12167

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, 128, 3757 – 3768. <https://doi.org/10.1121/1.3506351>

Lengeris, A., & Nicolaidis, K. (2014). Effect of phonetic training on the perception of English consonants by Greek speakers in quiet and noise. *Proceedings of Meetings on Acoustics*, 22(1), 1–6. https://doi.org/10.1121/2.0000025

Leong, C. X. R., Price, J. M., Pitchford, N. J., & Heuven, W. J. B. van. (2018). High variability phonetic training in adaptive adverse conditions is rapid, effective, and sustained. *PLOS* *ONE*, *13*(10), e0204888. https://doi.org/10.1371/journal.pone.0204888

Lerdpaisalwong, S. (2015). *Perception training of Thai learners: American English consonants and vowels* (Unpublished doctoral dissertation). The University of Wisconsin-Milwaukee.

Li, Y. (2015). *Audio-visual training effect on L2 perception and production of English /θ/-/s/ and /ð/-/z/ by Mandarin speakers* [Unpublished doctoral dissertation]. Newcastle University.

Li, Y. (2016). *Effects of high variability phonetic training on monosyllabic and disyllabic Mandarin Chinese tones for L2 Chinese learners* (Unpublished doctoral thesis). University of Kansas.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255. <https://doi.org/10.1121/1.408177>

Lively, S.E., Pisoni, D.B., & Yamada, R.A. (1994). Training Japanese listeners to identify English /r/ and/l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, 96, 2076 – 2087. <https://doi.org/10.1121/1.410149>

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. <https://doi.org/10.1121/1.1894649>

Lopez-Soto, T., & Kewley-Port, D., (2009). Relation of perception training to production of codas in English as a second language. *Proceedings of Meetings on Acoustics*, 6, 1 – 15. <https://doi.org/10.1121/1.3262006>

Macdonald, R.M.M. (2012). *Counteracting age related effects in L2 acquisition: Training to distinguish between French vowels* (Unpublished doctoral dissertation). The University of Edinburgh.

Melnik, G.A., & Peperkamp, S. (2020). High-variability phonetic training enhances second language lexical processing: Evidence from online training of French learners of English. *Bilingualism: Language and Cognition*, 24(3), 497 – 506. https://doi.org/10.1017/S1366728920000644

Mokari, P.G., & Werner, S. (2017). Perceptual training of second-language vowels: Does musical ability play a role? *Journal of Psycholinguistic Research*, 47, 95 – 112. <https://doi.org/10.1007/s10936-017-9517-8>

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1496–1509. <https://doi.org/10.1044/1092-4388>

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research*, *51*(6), 1480–1493. https://doi.org/10.1044/1092-4388(2008/07-0109)

Okuno, T., & Hardison, D.M. (2016). Perception-production link in L2 Japanese vowel duration: Training with technology. *Language Learning & Technology*, 20(2), 61 – 80. Retrieved from <http://llt.msu.edu/issues/june2016/okunohardison.pdf>

Pederson, E., Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America*, 127, EL54 – EL59. <https://doi.org/10.1121/1.3292286>

Protopapas, A., & Calhoun, B., (2000). Adaptive phonetic training for second language learners. *Proceedings of the 2nd International Workshop on Integrating Speech Technology in Language Learning* (pp. 31 – 38). UK: University of Abertay Dundee.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America*, *119*(3), 1684–1696. <https://doi.org/10.1121/1.2161427>

Qian, M. (2018). *An adaptive computational system for automated, learner-customized segmental perception training in words and sentences: Design, implementation, assessment* [Unpublished doctoral dissertation]. Iowa State University.

Qian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning & Technology*, 22(1), 69 – 96. <https://doi.org/10125/44582>

Rato, A. (2014). Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. *Proceedings of the International Symposium on the Acquisition of Second Language Speech Concordia Working Papers in Applied Linguistics*, 5, 529 – 546.

Reyes, Y.P., & Hazan, V., (2021). English vowel perception by non-native speakers: Impact of audio and visual training modalities. Onomazein: Journal of Linguistics, Philology and Translation, 51, 111 – 136. <https://doi.org/10.7764/onomazein.51.04>

Ryu, N.Y., & Kang, Y. (2019). Web-based high variability phonetic training on L2 coda identification. *Proceedings of the 19th International Congress of Phonetic Sciences*, 482 – 486.

Shinohara, Y. (2021). Audiovisual English /r/−/l/ identification training for Japanese-speaking adults and children. *Journal of Speech, Language, and Hearing Research*, *64*(7), 2529–2538. https://doi.org/10.1044/2021\_JSLHR-20-00506

Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *Journal of Phonetics*, *66*, 242–251. <https://doi.org/10.1016/j.wocn.2017.11.002>

Shinohara, Y., & Iverson, P. (2021). The effect of age on English /r/-/l/ perceptual training outcomes for Japanese speakers. *Journal of Phonetics*, *89*, 101108. <https://doi.org/10.1016/j.wocn.2021.101108>

Shport, I.A. (2016). Training English listeners to identify pitch-accent patterns in Tokyo Japanese. Studies in Second Language Acquisition, 38, 739 – 769. <https://doi.org/10.1017/S027226311500039X>

Shum, K.K.M., Au, T.K.F., Romo, L.F., & Jun, S.A. (2021). Learning challenging L2 sounds via computer training: High-variability perceptual training for children and adults. *Language Learning and Development*, 327 – 342. <https://doi.org/10.1080/15475441.2021.1876699>

Silpachai, A. (2020). The role of talker variability in the perceptual learning of Mandarin tones by American English listeners. *Journal of Second Language Pronunciation*, *6*(2), 209–235. <https://doi.org/10.1075/jslp.19010.sil>

Tajima, K., Kato, H., & Rothwell, A. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *The Journal of the Acoustical Society of America*, 123, 397 – 413. <https://doi.org/10.1121/1.2804942>

Thomson, R. I. (2012). Improving L2 listeners’ perception of English vowels: A computer-mediated approach. *Language Learning*, *62*(4), 1231–1258. https://doi.org/10.1111/j.1467-9922.2012.00724.x

Thomson, R. I. (2016). Does training to perceive L2 English vowels in one phonetic context transfer to other phonetic contexts? *Canadian Acoustics - Acoustique Canadienne*, *44*, 198–199.

Trakantalerngsak, T. (2016). *The effect of perceptual training on the learning of Japanese fricative and affricate contrasts by native Thai learners of Japanese* (Unpublished doctoral dissertation). Osaka University.

Ueda, R., & Hashimoto, K. (2019). Perceptual training in a classroom setting: Phonemic category formation by Japanese EFL learners. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference* (pp. 237–249). Iowa State University.

Wei, P. (2016). *Cross-linguistic perception and learning of Mandarin Chinese sounds by Japanese adult learners* (Unpublished doctoral dissertation). University of Oregon.

Wong, W.S. (2013). *Training the perception and production of English vowels /I/-/i:/, /e/-/æ/ and /ʊ/-/u/ by Cantonese ESL learners in Hong Kong* (Unpublished doctoral dissertation). The Chinese University of Hong Kong.

Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal*, *97*(1), 144–160. <https://doi.org/10.1111/j.1540-4781.2013.01386.x>

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, 32(4), 539–552. https://doi.org/10.1016/j.system.2004.09.011

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, 106(6), 3649–3658. https://doi.org/10.1121/1.428217

Yang, R., Nanjo, H., & Dantsuji, M. (2021). Self adaptive phonetic training for Mandarin nasal codas. *Computer-Assisted Language Learning Electronic Journal*, *22*(1), 391–413.

Yeon, S.-H. (2004). *Teaching English word-final alveolopalatals to native speakers of Korean* [Unpublished doctoral dissertation]. University of Florida.

Zhao, T.C., & Kuhl, P.K. (2015). Effect of musical experience on learning lexical tone categories, 137, 1452 – 1463. <https://doi.org/10.1121/1.4913457>