Mizumoto, A., Urano, K., & Maeda, H. (2014). A systematic review of published articles in ARELE 1–24: Focusing on their themes, methods, and outcomes. ARELE (Annual Review of English Language Education in Japan), 25, 33–48.

A Systematic Review of Published Articles in *ARELE* 1–24: Focusing on Their Themes, Methods, and Outcomes

Atsushi MIZUMOTO
Kansai University
Ken URANO
Hokkai-Gakuen University
Hiroaki MAEDA
Institute for Foreign Language Research and Education, Hiroshima University

Abstract

This study reviewed three representative aspects (themes, methods, and outcomes) of the published articles in *ARELE*, volumes 1 to 24. The review of 450 *ARELE* articles revealed the following results: (a) the 24 *ARELE* volumes could be divided into two periods (the first 12 and the second 12); (b) articles in each period have characteristic words to represent the themes peculiar to the period; (c) research themes have shifted from teaching to learning, with reading, vocabulary, assessment/testing, and motivation coming to the forefront; (d) articles are predominantly empirical studies, targeting learners at secondary and tertiary levels, and hypothesis generating, with a quantitative approach, while intervention studies are not common; (e) medium strength of effect size was obtained with a meta-analytical approach; (f) the effect size decreases toward more recent volumes, which may be a sign of theoretical refinement; (g) the statistical power of most studies is lower than it should be. A number of suggestions are offered for improving the quality of future research practice.

1. Introduction

Since 1990, *ARELE* (Annual Review of English Language Education in Japan) has been published annually by JASELE (the Japan Society of English Language Education), formerly known as FELES (the Federation of English Language Education Societies in Japan). The current issue marks the 25th anniversary of *ARELE*. *ARELE* claims that it is "one of the top journals in the field of English language education in Japan" (http://www.jasele.jp/ARELE/). With publication of the 25th volume of *ARELE*, it would be worthwhile undertaking a review study to understand the current field of research and its future direction.

A review study, as indicated by the ever-increasing number of meta-analyses (Plonsky & Oswald, 2012), plays a significant role in a research field that has expanded in scope and reached

relative theoretical maturity, because a review study is "a well-established way to broaden and deepen understanding in a field" (Stapleton, 2013, p. 145).

In fact, review studies of specific journals have increasingly been conducted in recent years (e.g., Hirano, 2011; Magnan, 2006; Stapleton & Collett, 2010; Terasawa, 2010). In particular, Stapleton and Collett (2010) reviewed 297 articles published in *JALT Journal* (the journal of the Japan Association for Language Teaching) over 30 years and revealed the changing trends in English language education in Japan.

Academic journals have begun reflecting on their research and practice, and *ARELE* is not an exception. Yanase (2013) checked the frequency of keywords often used in qualitative studies (i.e., reflective, reflection, qualitative, sociocultural, epistemological, and epistemology) in the titles and abstracts of articles published in *ARELE* volumes 13–22 (2002–2011). He then compared the use of keywords with that in *JACET Journal* and *Applied Linguistics* during the same period and found that *ARELE* published very few qualitative studies.

Yanase's study (2013) has certainly shed light on one element lacking in *ARELE* articles (i.e., qualitative studies). However, its focus was limited to one aspect of research methodology. In order to reveal a more detailed picture of the overall trends in the journal content over the years, we will investigate *ARELE* articles from a wider perspective. Thus, the primary purpose of this study is to address the following questions:

1. What research themes are studied in *ARELE*, and have they changed over time?
2. What types of research methods, targeting whom and for what purpose, are used?
3. What are the research outcomes (i.e., effect sizes) obtained from the studies?

2. Method

2.1 Article Collection

In volumes 1 to 24, a total of 479 articles have been published in *ARELE*. Of those 479 papers, 450 have been made available publicly at CiNii, an online database maintained by the National Institute of Informatics. In this study, therefore, those 450 articles were retrieved and analyzed. A corpus of 79,743 words (5,152 types) was then created using the titles and abstracts of the 450 articles.

2.2 Coding

In order to investigate the themes, we utilized the classification scheme used by JASELE's National Conference. It included the following 14 categories: assessment/testing, grammar, language policy, learner development/strategies, listening, materials, motivation, reading, second language acquisition/psycholinguistics, speaking, teacher development, teaching methodology, vocabulary, and writing. Of course, some studies could be classified

into more than one category; for example, “listening test” could be regarded as either listening or assessment/testing. In such a case, only one theme category was selected (e.g., “listening test” was counted as “listening”).

In addition to the basic information of the studies, such as study identification, publication year, volume number, title, author, we coded several features of research methodology for all the 450 articles (see Table 1). These methodological features were selected based on previous review studies (Hirano, 2011; Stapleton & Collett, 2010; Urano, 2012) for comparisons with the results of other review studies. Expectedly, some studies targeted more than one school level (e.g., comparing the test scores of junior high school students and those of university students). Again, in such a case, only one school level, the school level with younger learners, was selected (the case mentioned above was counted as “junior high school”).

The two types of coding described above (i.e., themes and research methods) were conducted by the first author of this paper and checked by the other authors. Inter-reliability of the coding was not checked because the purpose was to obtain a general picture of the themes and research methods.

Table 1
Features of Research Methodology and Their Categories

Features	Categories
Research type	Empirical study, Practical report, Survey report, Theoretical study
Target school level	Preschool, Elementary, Junior high, Senior high, University, Other
Research purpose	Hypothesis generating, Hypothesis testing, Other
Data type	Qualitative, Quantitative, Mixed, Other
Intervention	Present, Absent

Next, to evaluate the research outcomes (i.e., effect sizes), a meta-analytical approach was adopted to synthesize effect sizes obtained from a variety of studies in *ARELE*. Because we focused only on articles in *ARELE*, our meta-analysis was different from an ordinary one, in which effect sizes are synthesized across as many studies as possible, regardless of whether a paper has been published (Plonsky & Oswald, 2012). However, considering that meta-analysis offers the advantage of obtaining synthesized effect sizes across primary studies, it would be useful to pursue this approach for reviewing the outcomes of studies reported in *ARELE* to date.

In this study, we set the following inclusion criteria and coded the 450 articles accordingly:

- The article reports numerical data (336 of 450 articles).
- The study compares the mean scores of two groups (120 of 336 articles).

- The study employs a (quasi-)experimental design with a treatment group and a contrast group (28 of 120 articles).

We excluded the studies that do not report standard deviations (*SD*), or only report *p* values (without means, standard deviations, or test statistics). From the 28 articles, we extracted 63 cases that compare the mean scores of a treatment group and a contrast group.[1] For those 63 cases, all the necessary information to conduct a meta-analysis (i.e., the number of participants, means, and standard deviations) was retrieved. Table 2 shows the breakdown of the dataset used for the meta-analysis; these categories were used for subsequent moderator analyses.

Coding for the meta-analysis was conducted by the first author and checked by another author. From the original pool of 450 articles, 45 (10.0%) were randomly selected and coded by another rater to check the inter-reliability of the coding. The percentage agreement was 93.3.

Table 2
Breakdown of School Types and Skills of the Meta-analyzed Cases

School types	Grammar	Listening	Reading	Speaking	Vocabulary	Total
Junior high	0	3	5	0	0	8
Senior high	10	1	7	1	3	22
University	2	3	5	1	22	33
Total	12	7	17	2	25	63

2.3 Analysis

The corpus, composed of the titles and abstracts of 450 *ARELE* articles, was analyzed using the word count tool of CasualConc version 1.9.7 (Imao, 2013) to investigate the research themes. The words were lemmatized using Someya's e-lemma file (Someya, 1998). By excluding the common function words and numbers, a word list with 300 most frequent words for the 24 volumes of *ARELE* was obtained, resulting in a 300 × 24 matrix. A hierarchical cluster analysis (the Ward method with the squared Euclidean distance technique) was employed to determine whether the 24 volumes of *ARELE* could be grouped by volume (i.e., year of publication).

After applying the cluster analysis, characteristic words for the grouped volumes were explored. The characteristic words were detected with "keyness," using the log-likelihood statistic, available in the word count tool of CasualConc (see Imao, 2013, for details). With this analysis, we attempted to examine the characteristic words (themes) appearing in *ARELE* articles both cumulatively and over time.

Frequency counts were obtained and displayed in figures for the investigation of the research themes, using the classification scheme of JASELE's National Conference. In the

figures, the percentages for the grouped volumes, based on the results of the cluster analysis, were presented in addition to the overall frequency counts of each category to show the trends of research themes over time.

We used the same approach to analyze the features of research methodology. Frequency counts of the features of research methodology for each category as well as the percentages for the grouped volumes are displayed in the figures to reveal the changes over time. All the data analyses and plotting of the results were carried out using R version 3.0.1 (R Core Team, 2013).

A meta-analytical approach to synthesize the effect sizes across primary studies was conducted using metafor: meta-analysis package for R version 1.9-1 (Viechtbauer, 2010). Using the metafor package, the weighed effect sizes (i.e., Hedges's *g*, an index of standardized mean differences), were calculated.[2] Moderator analyses were subsequently conducted using the categories listed in Table 2. A random-effect model was employed to estimate the meta-analytic mean and variance (see Borenstein, Hedges, Higgins, & Rothstein, 2009; Plonsky & Oswald, 2012, for a discussion of model choices in meta-analysis).

Plonsky and Gass (2011) argued and demonstrated that early research in a given area tends to yield larger effect sizes because it is "often characterized by strong manipulations that set out to determine whether an effect exists and thereby determine whether the claims of a particular and usually novel hypothesis merit further attention" (p. 329). After an effect is found, subsequent studies focus on its generalizability, resulting in refined research designs and a steady decrease in effect sizes over time (Plonsky & Oswald, 2012). We therefore examined whether a change in effect sizes over time could be observed with the published papers in *ARELE* as well.

Although the studies, or cases investigated in the meta-analysis, were limited in number, power analysis was also utilized to further rate the study quality of the published papers in *ARELE*. Using the same dataset, the achieved power of the selected studies (cases) was examined with the R package "pwr" version 1.1.1 (Champely, 2012). In addition, the optimal sample size for future studies utilizing the same research design (i.e., comparing means of a treatment group and a contrast group) was estimated with the effect size obtained from the meta-analysis.

3. Results and Discussion

3.1 Research Themes of *ARELE* Articles

The results of the cluster analysis are presented in Figure 1. As can be seen in the figure, the 24 volumes of *ARELE* can be divided into the first half (Volumes 1–12) and second half (Volumes 13–24).

Table 3 shows the results of the corpus analysis of the titles and abstracts of *ARELE* articles. It lists the 30 most frequent words across all volumes and 30 characteristic words

for the first and second halves detected with the keyness analysis. In the case of the 30 most frequent words for all volumes, aside from those words used frequently in academic papers—such as study, result, effect, show, and present—theme-related words seem to appear in the list (i.e., read, test, word, task, teach, comprehension, vocabulary, or process).

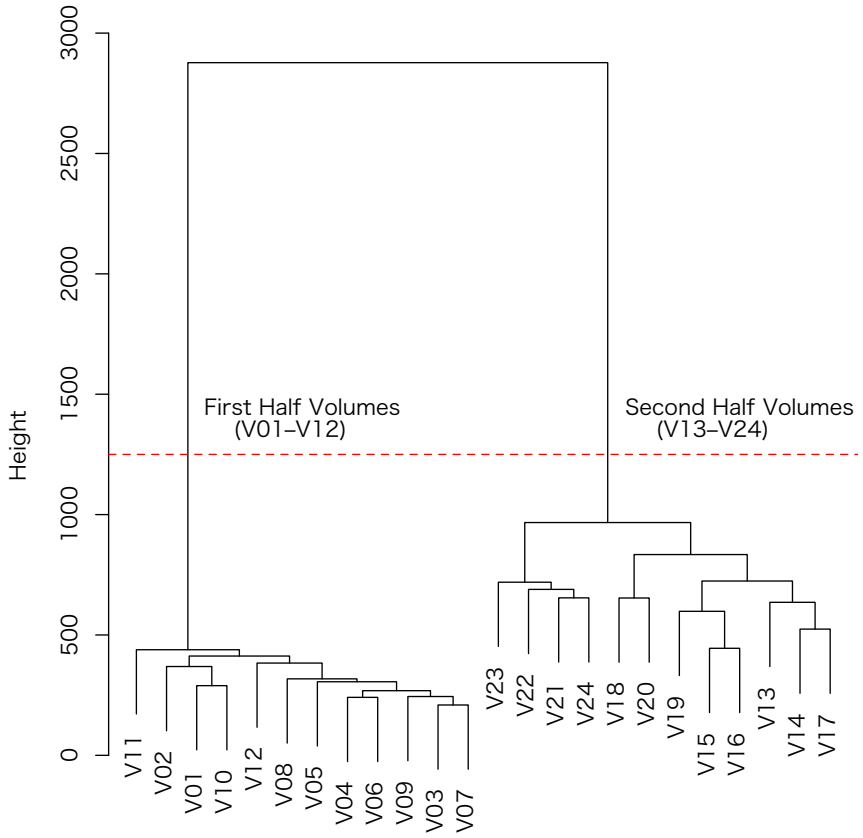

Figure 1. Result of cluster analysis (300 most frequent words × 24 volumes).

A comparison of the characteristic words for the first and second halves provides a more revealing picture of the trends in themes. The characteristic words for the first 12 volumes of *ARELE* suggest that themes related to communication, competence, (inter)cultural, team (as in "team teaching"), AETs, JTEs, passive voice, video, politeness, and syntax were more common in these 12 volumes.

On the other hand, the latter 12 volumes feature themes such as word, task, proficiency, vocabulary, performance, lexical, context, motivation, self, inference, repetition, peer, raters,

explicit, elementary, and assessment. Thus, we can assume that research topics including words such as motivation, self, peer (as in "peer assessment" or "peer feedback"), explicit (as in "explicit and implicit knowledge"), and elementary (as in "elementary school") have been attracting considerable attention among researchers and practitioners in the field of English language teaching in Japan in recent years.

Table 3
Most Frequent Words for All Volumes and Characteristic Words for the First and Second Halves

Volumes	Words
All (Volumes 1–24)	English, study, learner, Japanese, student, read, test, language, result, use, word, learn, high, school, EFL, task, effect, show, teach, write, comprehension, teacher, level, present, group, proficiency, L2, vocabulary, strategy, process
First half (Volumes 1–12)	language, teach, paper, communication, subject, competence, communicative, problem, production, cultural, cloze, team, intercultural, AETs, discussion, JTEs, passive, introduce, intonation, voice, video, ESL, emphasize, mechanism, American, politeness, syntax, consciousness, proposal, variability
Second half (Volumes 13–24)	study, read, word, learn, task, effect, proficiency, vocabulary, type, performance, lexical, item, context, motivation, target, participant, self, size, recall, scale, cue, inference, repetition, peer, sound, rater, pause, explicit, elementary, assessment

Note. For all volumes, the 30 most frequent words are listed in the frequency order. For the first (and second) half, these 30 words are more likely to occur in the first (and second) half compared with the other half.

Other striking differences between the first and second halves relate to how these two words are used: teach and learn. The word "teach" appears second in the list for the first 12 volumes, whereas the word "learn" appears fourth in the list for the latter 12 volumes. This may be due to the fact that the research interest of *ARELE* articles has shifted somewhat from teaching to learning. The other analysis of research themes (described below) corroborated this interpretation.

Figure 2 shows the results of the frequency counts using the classification scheme of the JASELE's National Conference. Teaching methodology is the most researched theme, but its frequency decreases in the latter 12 volumes. On the contrary, themes such as reading,

vocabulary, assessment/testing, and motivation increase in the latter volumes. In fact, if we merge these categories, other than teaching methodology, under the name "four skills" or "learning," a different picture emerges and the new category appears at the top of the list.

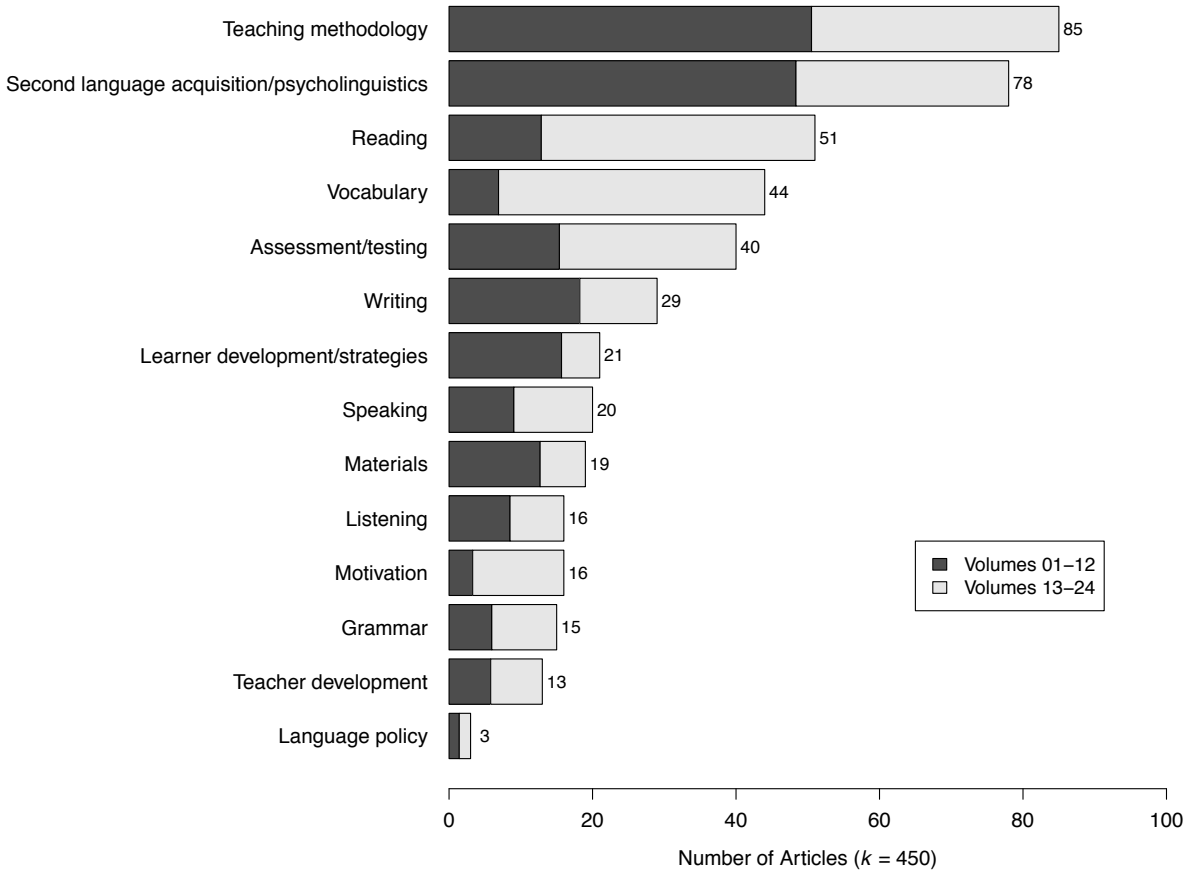

Figure 2. Research themes of the *ARELE* articles. The percentages for the first and second half of the 24 volumes are expressed in the bar plots in addition to the overall frequency counts to show the trends over time.

This finding is in line with the trends reported for the other academic journals, *JACET Journal* (Terasawa, 2010) and *JALT Journal* (Stapleton & Collett, 2010). Terasawa (2010) pointed out that articles in *JACET Journal* pay more attention to "the internal mechanism of learning" than to teaching English itself. In another study, Hirano (2011) reported similar trends in reviewing *CELES Journal* (even though teaching methodology accounts for more than 40 percent of all the articles). She attributed the increase in the number of articles on

reading and vocabulary to the theoretical and practical refinements in these research fields. The results of our study revealed similar trends in research themes in *ARELE* over time. That is, research on teaching methodology has possibly been marginalized to some extent as a result of theoretical advancement in other increasingly researched topic areas.

3.2 Research Methods of *ARELE* Articles

The results of the research method analysis of *ARELE* articles are presented in Figures 3 to 7. For comparison purposes, in addition to the overall frequency counts, the percentages of the first and second half of the 24 volumes are expressed in the bar plots to show the trends over time.

Figure 3 shows the research types of *ARELE* articles. Of all the articles, 74.4% (335 pieces) were empirical studies, with a higher number in the latter 12 volumes. A similar trend has been observed in the case of empirical studies in *JALT Journal* (Stapleton & Collett, 2010). Although *ARELE* has a specific section for practical reports (as of volume 24), only 8.2% articles accounted for this type of research. Considering the fact that one-fourth of the articles in *CELES Journal* are practical reports (Hirano, 2011), the number of practical reports in *ARELE* is very small.

Figure 4 illustrates the number and proportion of target school levels in *ARELE* articles. Studies conducted at the university level are by far the largest of all types, and they have been increasing toward the second half of the 24 volumes. This is partly because many researchers and practitioners working and conducting research at the university level submit their papers to *ARELE*.

Figure 5 presents the research purposes of *ARELE* articles. Theoretical studies were not included in this analysis. Of all the articles, 79.7% focused on hypothesis generating, whereas only 14.0% dealt with hypothesis testing. This high ratio of hypothesis generating studies does not necessarily imply that the field of English language education in Japan is still in the exploratory stage. Rather, we would like to claim that some of those hypothesis generating studies could have actually tested certain hypotheses if the authors had tried to narrow down their research questions by carefully examining existing literature.

Figure 6 indicates that the data types of *ARELE* articles are predominantly quantitative in nature (82.6%). Quantitative studies are also prominent in other journals, and they always outnumber qualitative inquiries in the field of applied linguistics (e.g., Magnan, 2006; Stapleton & Collett, 2010). Nevertheless, the ratio of quantitative studies in *ARELE* is noticeably larger than that in other journals, which is reflected in the very small proportion of qualitative studies (3.1%). One encouraging finding is that mixed-method studies have been increasing in the latter 12 volumes. Recently, Yanase (2013) called for support for reflective practice (i.e., qualitative inquiries). We would also argue that this finding clearly demonstrates the urgent need for a

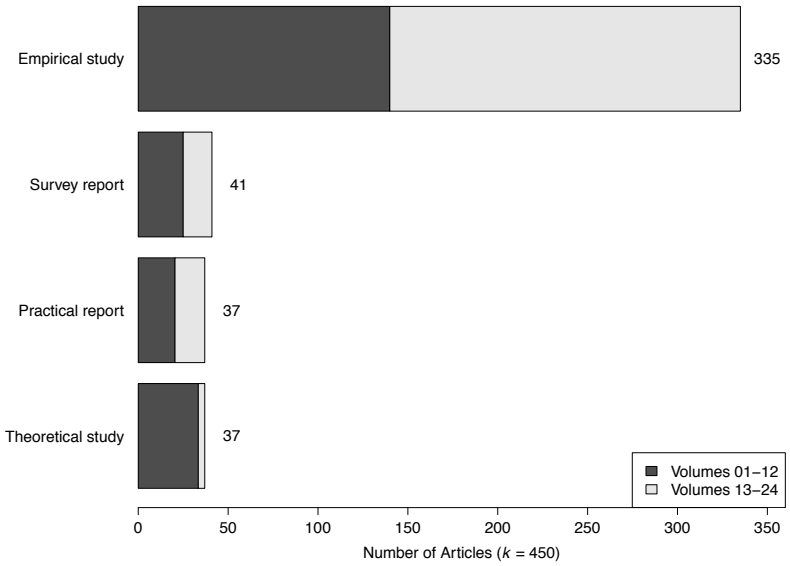

Figure 3. Research types of *ARELE* articles.

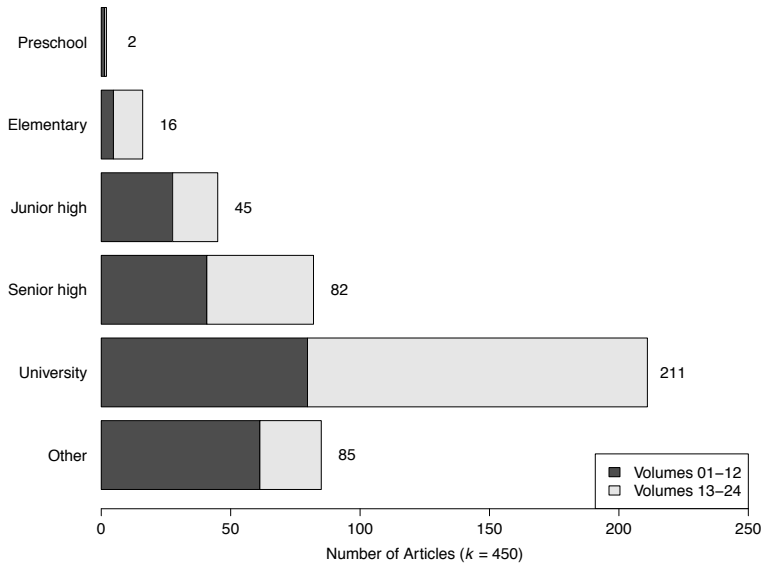

Figure 4. Target school levels researched in *ARELE* articles.

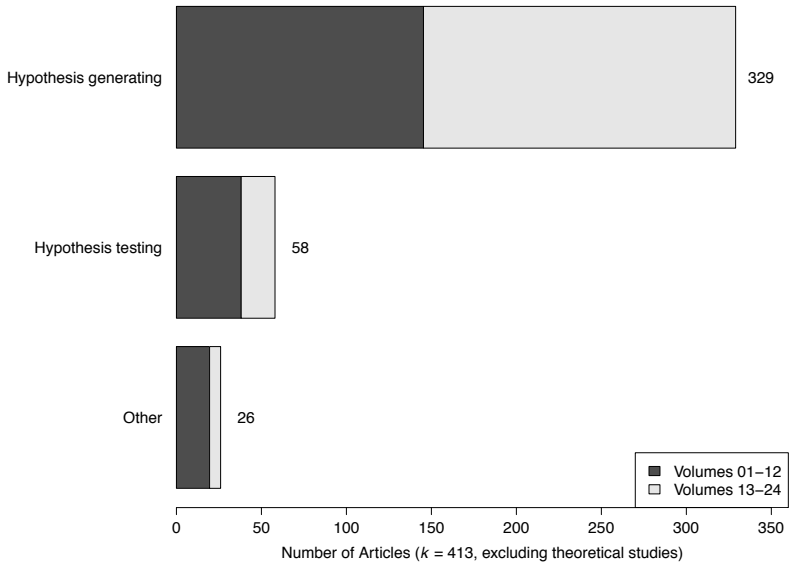

Figure 5. Research purposes of *ARELE* articles.

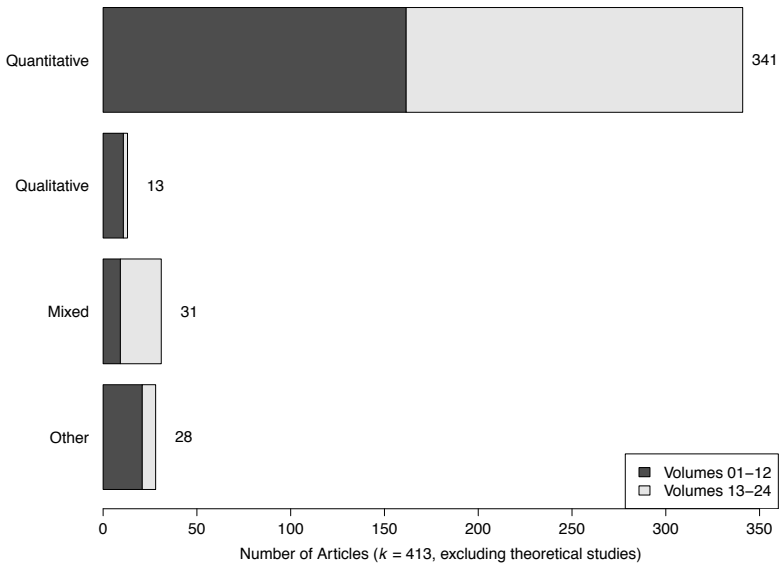

Figure 6. Data types of *ARELE* articles.

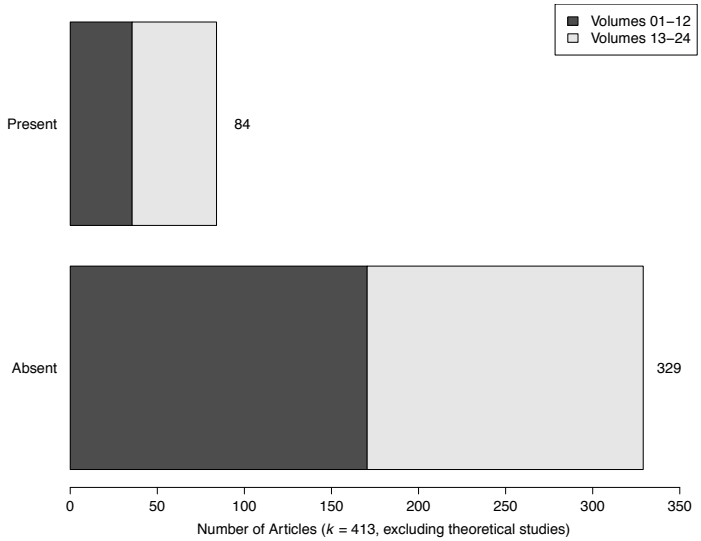

Figure 7. Intervention studies in *ARELE* articles.

more qualitative approach in dealing with diverse topics within the domains of English language education.

Figure 7 depicts that intervention studies were employed in one-fifth of the articles (20.4%). The number of intervention studies has been increasing in the latter 12 volumes, which is certainly a healthy sign in our view, because *ARELE* articles should strike a balance between research and practice in English language teaching.

3.3 Research Outcomes of *ARELE* Articles

Table 4 summarizes the results of the meta-analysis. The overall effect size (Hedges's g) according to the random-effect model was 0.76, 95% CI [0.59, 0.93]. Cohen (1988) proposed benchmarks for standardized mean differences, d (i.e., d = 0.20, small; d = 0.50, medium; d = 0.80, large). However, these standards are field-specific and Cohen himself argued they should be applied with caution. Recently, Oswald and Plonsky (2010), summarizing 27 meta-analyses of second language acquisition research, suggested the preliminary benchmarks: d = 0.40, small; d = 0.70, medium; d = 1.00, large. Interpreting the magnitude of effect size according to Oswald and Plonsky's benchmarks, the synthesized effect size from the *ARELE* articles was found to be of medium strength.

Because we did not focus on one research topic as in an ordinary meta-analysis, as expected, the effect varied considerably from one case to the next (Q = 396.05, df = 62, p < .001, I^2 = 87.22). Subsequent moderator analyses shown in Table 4 revealed more detailed

information about the differences in effect sizes across school types and skills. As for the school types, medium to large effect sizes were observed across three levels. With regard to skills, different degrees of effect sizes were found, with grammar having the smallest effect size (g = 0.38) and listening the largest (g = 1.49).

Table 4
Results of Meta-analysis (Random-effect Model)

Subgroup	k	Contrast group (n)	Treatment group (n)	Effect size (Hedges's g)	95% CI
Overall	63	3474	2476	0.76	[0.59, 0.93]
School type					
Junior high	8	347	479	0.89	[0.43, 1.35]
Senior high	22	1360	806	0.69	[0.40, 0.98]
University	33	1767	1141	0.77	[0.54, 1.01]
Skill					
Grammar	12	495	452	0.38	[0.04, 0.71]
Listening	7	687	237	1.49	[1.02, 1.96]
Reading	17	1182	794	0.91	[0.62, 1.21]
Speaking	2	30	30	0.88	[-0.04, 1.81]
Vocabulary	25	1080	913	0.64	[0.40, 0.88]

The meta-analysis conducted in the current study does not necessarily provide a conclusive answer to the impact of a specific research endeavor. However, it does present a telling case that, on average, the research articles in *ARELE*, which employ a (quasi-)experimental design with a treatment group and a contrast group, have positive effects to a certain degree (i.e., medium effect size).

Figure 8 shows the results of the change in effect sizes over time. Although different skills were included in the analysis and only cases on reading research were found in the 1990s, we can see a steady decrease in effect sizes. This result partly provides evidence that, as Plonsky and Gass (2011) claimed, theoretical progress has been made in the field of research and this is evident in the case of the *ARELE* articles as well. It should be borne in mind that given the limited number of studies and targeted specific research design, more research is necessary to confirm (or reject) this hypothesis.

The results of the power analysis showed that 27 cases out of 63 (42.9%) had reached enough power (> .80) based on Cohen's criterion (1988). In other words, 57.1% of the cases did not reach the power of .80. More seriously, 25 cases (39.6%) had a power less than .50 (i.e., worse than coin flipping). These results indicate that, although the reported effect sizes of

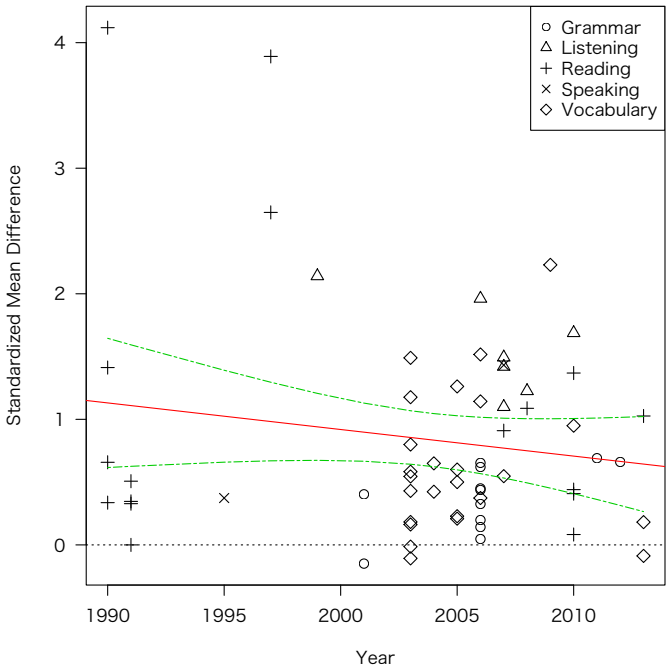

Figure 8. Publication years and effect sizes.

ARELE articles are of medium strength in general, statistical power is lower than it should be because of small sample sizes.

Finally, we calculated the optimal sample size for future studies, utilizing the same research design (i.e., comparing means of a treatment group and a contrast group) with the criteria of power = .80, α = .50 and effect size *d* (Hedges's *g* gained from meta-analysis) = 0.76. We found that 29 participants (n = 28.17 in the power analysis) would be necessary for each group for the same research design. Statistical power refers to the likelihood of finding a statistically significant result given that a difference really exists. If statistical power is low, the study cannot be replicated consistently (Schmidt & Hunter, 1997). That is, *p* values reported in one study will not be reproducible in others. Therefore, the power analysis shown here emphasizes the need for researchers to estimate and determine sample sizes prior to conducting their research project.

4. Conclusion

The purpose of this study was to understand the past and current state of research in English language education in Japan through a retrospective review of 450 articles published

in *ARELE*, one of the major journals in the field. We found a clear shift in the research themes investigated in the *ARELE* articles, that is, from teaching to learning, between the first and second halves of the past 24 volumes. We also pointed out that empirical studies and quantitative, as opposed to qualitative, research methods were overrepresented in *ARELE*. On a positive note, the effect sizes of the selected *ARELE* articles were of medium strength and they were decreasing toward more recent volumes, which may be a sign of theoretical refinement.

Because the characteristics reported in this paper are more or less in line with those of other domestic and international journals, we can state that, although some improvements are urgently needed, *ARELE* has led and shaped the research and practice of English language education in Japan. We do hope the next 25 volumes of *ARELE* will also help contribute to the progress of our profession.

Notes

1. Among those 63 cases, some cases included the comparison of the same treatment group and different contrast groups. That is, the same participants could be included in the cases as long as the study design had a treatment group and a contrast group.
2. Hedges's g is often interchangeably called d to refer to the same index in the literature. Borenstein et al. (2009) describe the differences in detail.

Acknowledgments

The authors would like to thank Dr. Yo In'nami and Dr. Rie Koizumi for their constructive comments and suggestions on meta-analysis.

References

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, U.K.: John Wiley & Sons. doi: 10.1002/9780470743386

Champely, S. (2012). pwr: Basic functions for power analysis (R package version 1.1.1) [Computer program]. Retrieved from http://cran.r-project.org/package=pwr

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Hirano, K. (2011). Hon gakkai ni okeru kenkyuu (1991-2010) no review to tenbou [An overview of English education research in CELES (1991-2010) and its prospects]. *CELES Journal*, *40*, 307–314.

Imao, Y. (2013). CasualConc (Version 1.9.7) [Computer software]. Retrieved from https://sites.google.com/site/casualconcj/

Magnan, S. S. (2006). From the editor: The MLJ turns 90 in a digital age. *The Modern Language Journal*, *90*, 1–5. doi: 10.1111/j.1540-4781.2006.00380.x

Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics*, *30*, 85–110. doi: 10.1017/S0267190510000115

Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning*, *61*, 325–366. doi: 10.1111/j.1467-9922.2011.00640.x

Plonsky, L., & Oswald, F. L. (2012). How to do a meta-analysis. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 275–295). London: Basil Blackwell. doi: 10.1002/9781444347340.ch14

R Core Team. (2013). R: A language and environment for statistical computing (Version 3.0.1) [Computer software]. Vienna, Austria. Retrieved from http://www.r-project.org/

Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. L. E. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 37–64). Mahwah, NJ: Lawrence Erlbaum Associates.

Someya, Y. (1998). e_lemma.txt (Version 2 for WordSmith 4) [Word list]. Retrieved from http://www.lexically.net/downloads/version4/e_lemma.zip

Stapleton, P. (2013). Using conference submission data to uncover broad trends in language teaching: A case study of one conference over 30 years. *Language Teaching Research*, *17*, 144–163. doi: 10.1177/1362168812460808

Stapleton, P., & Collett, P. (2010). JALT Journal turns 30: A retrospective look at the first three decades. *JALT Journal*, *32*, 75–90. Retrieved from http://jalt-publications.org/files/pdf-article/perspectives_0.pdf

Terasawa, T. (2010). *Kyouiku kenkyuu to shite no gaikokugo kyouikugaku* [Foreign language education as educational research]. Retrieved from http://d.hatena.ne.jp/TerasawaT/20101102/1288721439

Urano, K. (2012). *Kiyou ronbun no bunseki* [An analysis of CELES Journal articles]. Retrieved from http://www.urano-ken.com/research/project/project_3.pdf

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. Retrieved from http://www.jstatsoft.org/v36/i03/

Yanase, Y. (2013). *Reflective na eigo kyouiku: 10 nen kan no doukou* [Reflective English language education: A trend for a decade]. Retrieved from http://yanaseyosuke.blogspot.jp/2013/08/10.html