OPEN ACCESS
Future Education Review (FER)
Vol. 1 No. 1. May 2026, pp. 31-37
DOI. https://doi.org/10.53893/fer.v1i1.511

Research Article

BEYOND DRILLS: THE EFFECTS OF DEVELOPMENTALLY ALIGNED VOCABULARY GAMES ON EFL SPEAKING AMONG TAIWANESE ELEMENTARY LEARNERS

Ching-Wen Kao
Curriculum and Instruction Program, Department of Education, National University of Tainan, Taiwan

Corresponding author email: angel@gnpro.org

Article Info
Recieved: 1/28/2026
Revised: 2/22/2026
Accepted: 4/29/2026
OnlineVersion: 5/01/2026
Check for updates

Abstract
This quasi-experimental study investigated whether English vocabulary games aligned with Piaget's concrete operational stage could improve the EFL speaking proficiency of elementary learners in a Taiwanese after-school setting. Twenty Grade 3–4 students (ages 8–10) were assigned to an experimental group ($n = 10$) receiving a 16-week game-based intervention or a control group ($n = 10$) continuing with conventional drills. Speaking-related performance—self-efficacy, vocabulary use, and fluency—was measured via a Likert-scale questionnaire (Cronbach's $\alpha = .85$). The experimental group showed large, statistically significant gains across all subscales, $t\,(9) = 8.72$, $p < .001$, $d = 1.45$, while the control group showed no change. Between-group comparison confirmed a substantial posttest advantage, $t\,(18) = 8.66$, $p < .001$, $d = 3.87$. Qualitative responses corroborate the quantitative patterns. These findings suggest that developmentally aligned game design offers a theoretically grounded approach to supporting young EFL learners' speaking development.

Keywords: gamification, EFL speaking, Piaget's theory, cognitive development, concrete operational stage, elementary education, quasi-experimental design, Taiwan

INTRODUCTION

English speaking remains one of the most challenging skills for elementary learners in Taiwan, where classroom instruction is historically weighted toward grammar and reading. This imbalance limits students' opportunities for oral practice and can erode both self-efficacy and willingness to communicate—two constructs widely regarded as proximal determinants of speaking achievement (Zhang & Hasim, 2023). Instructional approaches that create low-stakes, engaging speaking environments are therefore worth systematic investigation.

Gamified instruction has drawn growing empirical interest as one such approach. Research documents that game elements—feedback loops, incremental challenge, and social competition—can activate intrinsic motivation and sustain engagement across diverse educational contexts (Hamari et al., 2014; Sailer et al., 2017). Within EFL settings, gamification has been associated with gains in vocabulary retention, fluency, and positive attitudes toward the target language (Zhang & Hasim, 2023). However, the extant literature largely treats game design as a motivational overlay rather than as a vehicle for developmentally calibrated cognition, leaving the theoretical rationale connecting game mechanics to young learners' cognitive demands underspecified.

Piaget's (1952) theory of cognitive development provides a compelling framework for closing this gap. Children aged 7–11 operate primarily within the concrete operational stage, during which learning is most effective when tasks engage classification, matching, and logical sequencing with tangible materials. Designing vocabulary games that mirror these operations should reduce the cognitive burden of new linguistic forms and support deeper encoding. Despite the intuitive appeal of this synthesis, empirical studies that operationalise Piagetian stage theory in EFL game design and subject the resulting intervention to controlled evaluation are notably absent from the literature.

The present study addresses this gap by investigating whether a 16-week curriculum of Piaget-aligned vocabulary games could significantly improve the speaking-related performance—operationalised as self-efficacy, vocabulary use, and fluency—of elementary EFL learners in a Taiwanese after-school setting, relative to a control group receiving conventional instruction. In doing so, it responds to calls for theoretically grounded gamification research that moves beyond correlation toward causal inference (Dicheva et al., 2015).

LITERATURE REVIEW

Gamification and EFL SpeakingResearch Design

This Gamification—the application of game design elements in non-game contexts to foster motivation and behaviour change (Dicheva et al., 2015) has accumulated a substantial empirical base over the past decade. Hamari et al. (2014) synthesised 24 empirical studies and concluded that gamification generally produces positive motivational and behavioural outcomes, though effects are moderated by contextual variables and the specific mechanics deployed. In language education, Dicheva et al. (2015) mapped 34 published studies and identified feedback, points, and challenge as the most consistently effective game elements for sustaining engagement and improving learning outcomes.

Zhang and Hasim (2023) conducted a systematic review of 40 peer-reviewed EFL/ESL studies and found robust effects on vocabulary and speaking, with larger gains when game design was paired with instructional scaffolding. This finding is directly relevant to the teacher-guided format adopted in the present study. Sailer et al. (2017) demonstrated experimentally that specific elements—badges, leaderboards, and performance graphs—satisfy the psychological needs of competence and relatedness identified by self-determination theory as prerequisites for intrinsic motivation, providing a mechanistic account of why gamified environments tend to reduce speaking anxiety and increase willingness to communicate.

Clark et al. (2016) conducted a meta-analysis across K–16 settings ($k = 77$ comparisons) and reported a mean effect of $\overline{g} = 0.33$ for game versus non-game conditions, with effects amplified when task design incorporated explicit scaffolding. Although the evidence base for elementary EFL specifically remains geographically concentrated, the convergence of these findings supports the premise that well-designed gamification can improve both motivational and linguistic outcomes for young learners.

Piaget's Theory of Cognitive Development and Game-Based Learning

Jean Piaget's (1952) account of intellectual development posits four invariant stages, each characterised by qualitatively distinct cognitive structures. Two stages are of direct relevance here. During the preoperational stage (approximately 2–7 years), children exhibit symbolic function and enjoy pretend play, yet their thinking remains egocentric. During the concrete operational stage (approximately 7–11 years), logical operations become possible when grounded in tangible, observable materials: children can classify, seriate, and understand conservation. Piaget argued that these operations are not transmitted didactically but constructed through active manipulation of the environmental process he termed equilibration, the dynamic interplay between assimilation and accommodation.

Play occupies a central position in Piagetian theory as the vehicle through which children exercise and consolidate cognitive operations (Piaget, 1952). The rule-governed games of the concrete operational stage represent socially regulated contexts for practising logical operations—a characterisation that maps directly onto the structured game activities employed in the present intervention. Contemporary game-based learning research corroborates this: Clark et al. (2016) found that task alignment with cognitive scaffolding amplifies motivational effects, and Sailer et al. (2017) showed that game mechanics calibrated to competence level are associated with higher psychological need satisfaction.

Theoretical Gaps and the Present Study

Despite convergent interest in gamification and constructivist pedagogy, few studies have deliberately operationalised Piaget's stage theory as a design principle for EFL vocabulary games and evaluated the result with experimental controls (Dicheva et al., 2015; Zhang & Hasim, 2023). Studies in Taiwan have documented motivational benefits of vocabulary games at the elementary level, but most rely on single-group designs or cross-sectional surveys without direct speaking measures, limiting causal inference. The present study addresses three gaps: (a) the absence of theoretically guided, stage-aligned game design; (b) the underrepresentation of EFL speaking as a dependent variable in gamification research; and (c) the scarcity of controlled quasi-experimental designs in Taiwanese after-school EFL.

METHOD

Participants and Setting

Participants were 20 elementary school students (Grades 3 and 4; ages 8–10 years; 11 female, 9 male) enrolled in a voluntary after-school English club at a public primary school in Pingtung City, Taiwan. All participants had received school-based English instruction for one to two years prior to the study. Students were assigned to an experimental group (n = 10) or a control group (n = 10). Full random assignment was not feasible given the club's intake structure; however, baseline equivalence was confirmed statistically (see Results). Informed written consent was obtained from parents and guardians, and the study protocol was approved by the institutional review board of the affiliated university.

Design and Procedure

A two-group, pretest–posttest quasi-experimental design was employed across a single 16-week semester. Both groups meet once weekly for 50 minutes, with the first author serving as instructor for both, thereby eliminating instructor variability as a confound. The pretest was administered in Week 1, the intervention ran from Weeks 2 through 15, and the posttest was conducted in Week 16.

Experimental condition

The experimental group participated in vocabulary games explicitly designed to mobilise the cognitive operations of the concrete operational stage (see Table 1). Weekly vocabulary sets (e.g., tiger, apple, rice, book, tree, moon) were introduced through three formats: (a) a teacher-cued spelling task requiring perceptual classification before target-word production (e.g., "An orange striped animal? Spell: tiger"), (b) a 50-card memory-matching activity pairing visual stimuli with orthographic or phonological representations, and (c) a short oral practice round in which small groups (4–5 students) earned points for fluency and intelligibility. Game complexity progressed from concrete classification (Weeks 2–13) to integrative review (Weeks 14–15).

Control condition

The control group engaged in conventional drills—flashcard review, choral repetition, and written exercises—without game-based elements. Vocabulary coverage and weekly time-on-task were held constantly across conditions.

Table 1. 16-Week Piaget-Aligned Curriculum Overview

Week	Activity Focus	Vocabulary Set	Piagetian Operation
1	Pretest; game orientation	—	Symbolic function (preoperational)
2	Spelling challenge	Tiger	Classification
3	Memory matching (50 cards)	Apple	Conservation of correspondence
4–7	Oral practice rounds	Rice, Drink, Book, Pen	Logical sequencing

8–13	Vocabulary expansion	Tree, Flower, Sun, Animal, Food, Moon	Seriation; class inclusion
14–15	Integrative review	Book–review, Animal	Early abstract generalisation
16	Posttest	—	—

Note. Game complexity progressed from concrete classification tasks (Weeks 2–13) to integrative review (Weeks 14–15).

Instrument

Speaking-related performance was assessed using a researcher-adapted 5-item Likert-scale questionnaire (1 = strongly disagree, 5 = strongly agree) organised into three subscales: self-efficacy (e.g., "I feel confident when I speak English in class"), vocabulary use (e.g., "I can use new English words when I speak"), and fluency (e.g., "I can speak English without stopping too often"). Three experienced EFL educators reviewed the instrument for content validity and age-appropriateness. A pilot test with 12 separate students yielded Cronbach's $\alpha = .85$, indicating acceptable internal consistency (Dunn et al., 2014).

Data Analysis

Quantitative data were analysed using IBM SPSS Statistics Version 28.0 (George & Mallery, 2019), with supplementary effect-size and normality diagnostics via the Pingouin package for Python (Vallat, 2018). The significance threshold was set at $\alpha = .05$, two-tailed. Normality was confirmed via Shapiro–Wilk tests (all $W > .90$, $p > .05$). Cohen's d was the primary effect-size index (small ≥ .20, medium ≥ .50, large ≥ .80; Cohen, 1988). Bootstrapped 95% confidence intervals (5,000 resamples) around d guarded against capitalisation on chance. A priori power analysis in G*Power 3.1 (Faul et al., 2009) indicated n = 26 per group for d = 0.80, $\alpha = .05$, power = .95; the observed d = 1.45 yielded post-hoc power = 1.00. Open-ended responses were coded thematically following Braun and Clarke (2006).

RESULTS AND DISCUSSION

Results

Baseline Equivalence

The two groups were statistically equivalent at pretest on all subscales and the total score (M = 195.00 for both; SDexp = 4.24, SDctrl = 4.50), and an independent-samples t-test confirmed no significant pre-existing difference, $t(18) = 0.17$, $p = .87$.

Descriptive Statistics

Table 2 presents group means, standard deviations, and 95% confidence intervals by time point. The experimental group showed clear gains across all subscales from pretest to posttest, whereas the control group remained unchanged.

Table 2. Descriptive Statistics by Group and Time Point

Subscale	**Group**	**n**	**Pre M (SD)**	**95% CI**	**Post M (SD)**	**95% CI**	**Range**
Self-Efficacy	Experimental	10	35.0 (2.11)	[33.6, 36.4]	45.0 (1.49)	[44.0, 46.0]	30–50
	Control	10	35.0 (2.24)	[33.4, 36.6]	35.0 (2.24)	[33.4, 36.6]	30–50
Vocabulary Use	Experimental	10	37.0 (2.00)	[35.6, 38.4]	46.0 (1.58)	[45.0, 47.0]	32–50
	Control	10	37.0 (2.45)	[35.2, 38.8]	37.0 (2.45)	[35.2, 38.8]	32–50
Fluency	Experimental	10	41.0 (2.00)	[39.6, 42.4]	48.0 (1.83)	[46.9, 49.1]	36–50

	Control	10	41.0 (2.58)	[39.1, 42.9]	41.0 (2.58)	[39.1, 42.9]	36–50
Total	Experimental	10	195 (4.24)	—	235 (3.08)	—	180–250
	Control	10	195 (4.50)	—	195 (4.50)	—	180–250

Note. Subscale scores range from 10 to 50; total score from 30 to 150. CI = confidence interval.

Within-Group Inferential Statistics

Paired-samples t-tests for the experimental group revealed statistically significant gains on all measures (Table 3). Self-efficacy: $t(9) = 8.50$, $p < .001$, $d = 1.40$; vocabulary use: $t(9) = 7.75$, $p < .001$, $d = 1.28$; fluency: $t(9) = 7.21$, $p < .001$, $d = 1.19$; total: $t(9) = 8.72$, $p < .001$, $d = 1.45$ (bootstrapped 95% CI: [1.12, 1.78]). All gains exceeded the distribution-based minimal important difference of 0.5 × SDbaseline (Cohen, 1988). The control group showed no significant change on any measure.

Table 3. Paired-Samples t-Tests for the Experimental Group (df = 9)

Subscale	**t**	**p**	**MΔ (95% CI)**	**d**	**Power**
Self-Efficacy	8.50	< .001	10.0 [8.1, 11.9]	1.40	1.00
Vocabulary Use	7.75	< .001	9.0 [7.0, 11.0]	1.28	.99
Fluency	7.21	< .001	7.0 [5.0, 9.0]	1.19	.98
Total	**8.72**	**< .001**	**40.0 [32.1, 47.9]**	**1.45**	**1.00**

Between-Group Comparison

An independent-samples t-test on posttest total scores confirmed a large advantage for the experimental condition: $t(18) = 8.66$, $p < .001$, $d = 3.87$ (95% CI for MD: [32.1, 47.9]; experimental M = 235.0 vs. control M = 195.0).

Qualitative Findings

Thematic analysis (Braun & Clarke, 2006) of open-ended responses yielded two primary themes. Affective engagement was evident in 90% of experimental-group responses, characterised by language of enjoyment and reduced anxiety (e.g., "I like playing games," "I'm not scared to say it wrong"). Perceived competence emerged in approximately 80% of experimental responses (e.g., "I know more words now," "I can say more"). Control-group responses clustered around themes of routine and repetition. These patterns corroborate the quantitative findings and suggest that motivational and cognitive mechanisms were jointly activated by the intervention.

Discussion

The present study provides controlled evidence that a 16-week program of vocabulary games designed in accordance with the concrete operational stage can significantly improve elementary EFL learners' self-reported speaking proficiency. The experimental group's pre-to-post gains were large by conventional benchmarks (Cohen, 1988), remained stable under bootstrapped resampling, and were not replicated in the control groups supporting the interpretation that improvement was attributable to the intervention rather than to maturation or general instructional exposure.

These findings extend the gamification of literature in at least two substantive ways. First, they demonstrate that game-based instruction can improve speaking-related self-perceptions, not only vocabulary knowledge, the most studied outcome in this literature (Zhang & Hasim, 2023). Second, the study illustrates that the theoretical coherence of game design matters: aligning task demands with the classification, matching, and seriation operations characteristic of the concrete operational stage appears to have reduced the cognitive gap between what learners were asked to do linguistically and what they were developmentally prepared to do. This interpretation is consistent with Piaget's (1952) concept of equilibration and with evidence that motivational effects are amplified when challenge is calibrated to competence (Clark et al., 2016; Sailer et al., 2017).

The motivational mechanisms documented by Sailer et al. (2017)—satisfaction of competence and relatedness needs through feedback and social game structures—are likely to have operated alongside the developmental alignment described above. Students' qualitative responses, which emphasised enjoyment and confidence, are consistent with this dual-mechanism account and align well-established findings that low-stakes contexts can attenuate the affective filter inhibiting oral production in foreign language learners (Hamari et al., 2014).

Several limitations constrain generalisation. The sample of 20 students from a single after-school club is small, and the absence of full random assignment means unmeasured pre-existing differences cannot be entirely excluded despite statistical baseline equivalence. The self-report questionnaire captures perceptions of speaking ability rather than independently scored oral performance. Future studies should incorporate validated speaking rubrics or audio-recorded fluency measures, larger samples, and formal classroom settings to test the robustness and transferability of these findings.

Despite these limitations, the study carries practical implications. Short, structured vocabulary games—requiring no digital infrastructure—can meaningfully support speaking development when designed to engage the cognitive operations learners are developmentally positioned to apply. For curriculum designers, the findings illustrate a principled methodology for connecting developmental theory to game design: rather than adding gamification as a motivational overlay, practitioners might ask which game structures best instantiate the cognitive demands their learners are ready to meet (Dicheva et al., 2015).

CONCLUSION

This quasi-experimental study examined the effect of Piaget-aligned vocabulary games on the EFL speaking proficiency of elementary students in a Taiwanese after-school setting. A 16-week game-based intervention produced large, statistically significant gains in self-efficacy, vocabulary use, and fluency for the experimental group, with no corresponding change in the control group. The theoretical synthesis of Piagetian cognitive-developmental principles and gamification design offers a coherent basis for understanding why certain game structures may be more effective than others with young language learners. Future research scaling these findings to larger samples, more diverse instructional contexts, and objective measures of oral production would strengthen the evidence base for developmentally grounded game design in EFL education.

REFERENCES

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. Review of Educational Research, 86(1), 79–122.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. Educational Technology & Society, 18(3), 75–88.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology, 105(3), 399–412.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160.

George, D., & Mallery, P. (2019). IBM SPSS Statistics 26 step by step: A simple guide and reference (16th ed.). Routledge.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In Proceedings of the 47th Annual Hawaii International Conference on System Sciences (pp. 3025–3034). IEEE.

Piaget, J. (1952). The origins of intelligence in children. International Universities Press.

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. Computers in Human Behavior, 69, 371–380.

Vallat, R. (2018). Pingouin: Statistics in Python. Journal of Open Source Software, 3(31),

Zhang, S., & Hasim, Z. (2023). Gamification in EFL/ESL instruction: A systematic review of empirical research. Frontiers in Psychology, 13.