

Bilingual Screen Media and Peer Play in Preschool Language Development: A Mexico–Spain Comparison

Sofia Rodriguez^{1,*} and Javier Morales²

¹UNAM (National Autonomous University of Mexico), Mexico

²Universidad Politécnica de Madrid, Spain

E-mail: sofia.rodriguez@mock.iccmets.test; javier.morales@mock.iccmets.test

Abstract.

Building on prior findings on screen time and peer interaction, this cross-national study of 260 preschoolers in Mexico City and Madrid examines how structured bilingual screen content combined with guided peer play affects vocabulary growth in a second language. Earlier research indicated that peer interaction supports early second language vocabulary more effectively than unstructured screen exposure, but it left open whether screen media designed for language learning can add value when it is coupled with, rather than substituted for, interactive play. Children aged four to five years attending public preschools in the two cities were assigned by classroom to one of three twelve-week conditions: structured bilingual screen content alone, guided peer play alone, or a combined condition in which screen episodes introduced target vocabulary that was then practised in teacher-guided peer play. English was the target language in both cities. Receptive and expressive vocabulary were assessed before and after the intervention with picture-based measures. Multilevel models showed that the combined condition produced the largest gains in both receptive and expressive vocabulary, with effect sizes of 0.62 and 0.71 relative to the screen-only condition, and that guided peer play alone outperformed screen content alone on expressive vocabulary. Gains in the combined condition were of similar magnitude in Mexico City and Madrid, despite differences in home exposure to English. The findings suggest that bilingual screen media contribute to second language development when they are embedded in interactive play, and they offer guidance for preschool programmes seeking to use media purposefully.

Keywords: second language acquisition, preschool, bilingual screen media, peer play, cross-national comparison

**Corresponding author: sofia.rodriguez@mock.iccmets.test*

1. Introduction

Young children in Spanish-speaking countries are increasingly exposed to English through screen media before they receive any formal instruction. Streaming platforms, educational applications and broadcast programming offer bilingual and English-language content aimed at preschool audiences, and parents frequently regard such exposure as a foundation for later language learning. Research on early second language acquisition, however, has cautioned that passive screen exposure produces limited vocabulary gains in comparison with social interaction, and public health guidance continues to recommend limits on screen time for young children (World Health Organization, 2019). The question for early childhood educators is not whether screens should be present, since they are, but whether screen content can be used in ways that contribute to language development.

In our earlier study (Rodriguez, 2025) we compared the effects of screen time and peer interaction on second language acquisition in early childhood, using observational and assessment data from preschools in Mexico. That study found that time spent in peer

interaction involving the second language predicted vocabulary growth more strongly than time spent with second language screen content, and that screen exposure was associated with gains only in receptive vocabulary. It concluded that interaction is the primary driver of early second language learning but noted that the screen content children encountered was largely unstructured entertainment, and that the study could not determine whether content designed for language learning would behave differently, particularly if it were coupled with interactive activity rather than consumed alone.

The present study addresses that question through a cross-national quasi-experiment in Mexico City and Madrid. It compares three twelve-week conditions in public preschool classrooms: structured bilingual screen content alone, guided peer play alone, and a combined condition in which screen episodes introduce vocabulary that is then practised in teacher-guided peer play. The inclusion of two cities with different levels of ambient English exposure allows the study to examine whether effects depend on context. The target language is English, which is the second language most commonly introduced in preschools in both countries.

The study asks three questions. Does structured bilingual screen content produce second language vocabulary gains when used alone? Does combining screen content with guided peer play produce larger gains than either component alone? Are effects consistent across the two national contexts? The paper reviews relevant literature on early second language acquisition, screen media and play-based learning, describes the design and measures, reports results from multilevel analyses, discusses implications and concludes with limitations and directions for future research.

2. Literature Review

2.1 Early second language vocabulary acquisition

Vocabulary is the foundation of early second language development and the component most sensitive to input quantity and quality. Research on bilingual development shows that children's vocabulary in each language reflects their exposure to it, and that the interactive quality of exposure matters as much as its amount (Hoff et al., 2012). Sociocultural theory holds that language is acquired through participation in meaningful joint activity with more capable others, who scaffold the child's use of new forms (Vygotsky, 1978). For second language learning in preschool, joint activity with peers is especially important because it provides opportunities to produce the language, not only to hear it, and production is associated with expressive vocabulary growth. Studies of immersion and bilingual preschools consistently find that children who engage in more peer talk in the target language show greater expressive gains, and our previous research (Rodriguez, 2025) confirmed this pattern in Mexican settings. Paradis (2011) further shows that child-external factors such as richness of input interact with child-internal factors such as age and memory, so that interventions must be evaluated with attention to individual variation.

2.2 Screen media and language learning in early childhood

Evidence on screen media and early language is nuanced. Experimental work by Kuhl et al. (2003) showed that infants exposed to a foreign language through live interaction learned phonetic distinctions while infants exposed through video did not, a result often cited as the video deficit. Subsequent research has qualified this finding for older children: preschoolers can learn words from video, particularly when content is designed with repetition, labelling and opportunities for response, and when an adult co-views and interacts (Roseberry et al., 2014). Reviews conclude that educational content designed with developmental principles

produces modest vocabulary gains, while entertainment content does not (Madigan et al., 2020; Anderson & Subrahmanyam, 2017). The distinction between structured and unstructured content, and between solitary and co-viewed exposure, is therefore central. The present study uses structured bilingual content with explicit vocabulary targets and examines it both alone and embedded in interaction.

2.3 Guided play and the integration of media

Guided play occupies a position between free play and direct instruction, in which adults set up play environments and scaffold children's activity toward learning goals while children retain agency (Weisberg et al., 2016). Evidence indicates that guided play supports vocabulary learning more effectively than either free play or didactic instruction, because it combines the engagement of play with the focus of instruction (Toub et al., 2018). Integration of media into guided play has been proposed as a way of using screen content purposefully: a screen episode introduces vocabulary and a narrative context, and subsequent play provides opportunities to use the vocabulary with peers under adult scaffolding. This design aligns with the joint media engagement framework (Takeuchi & Stevens, 2011), which holds that media contribute to learning when they are embedded in social interaction. Drawing on this literature, the study hypothesises that the combined condition will produce larger gains in receptive and expressive vocabulary than either the screen-only or play-only condition (H1), that play-only will outperform screen-only on expressive vocabulary (H2), and that the pattern of effects will be consistent across the two cities (H3).

3. Methodology

3.1 Design and participants

A cluster quasi-experimental design was used, with preschool classrooms assigned to conditions. Twelve classrooms participated, six in public preschools in Mexico City and six in public preschools in Madrid, with two classrooms per condition in each city. Assignment was random within city, subject to the constraint that classrooms in the same school were assigned to different conditions to reduce contamination. Participants were 260 children aged four to five years whose parents consented, 132 in Mexico City and 128 in Madrid. Table 1 presents the sample profile by condition and city. Children with diagnosed language disorders or with English as a home language were excluded. Ethical approval was granted by UNAM and the Universidad Politécnica de Madrid, and by the participating education authorities.

Table 1: Sample profile by condition and city (N = 260)

Characteristic	Screen only	Play only	Combined	Total
Children, Mexico City	44	43	45	132
Children, Madrid	42	44	42	128
Total children	86	87	87	260
Mean age, months (SD)	56.3 (5.1)	56.8 (4.9)	56.1 (5.3)	56.4 (5.1)
Girls (%)	48.8	50.6	49.4	49.6
Home English exposure, Mexico City (%)	18.2	16.3	20.0	18.2
Home English exposure, Madrid (%)	35.7	36.4	33.3	35.2
Baseline receptive vocabulary, mean (SD)	9.4 (4.2)	9.7 (4.5)	9.2 (4.1)	9.4 (4.3)

Source: Authors' study data (2026)

3.2 Intervention conditions

All three conditions ran for twelve weeks with three sessions per week of approximately 30 minutes, delivered by the regular classroom teacher with support from a research assistant. The vocabulary targets were the same in all conditions: 72 English words drawn from themes of home, animals, food, body and daily routines, introduced at six words per week. In the screen-only condition, children watched a bilingual animated episode of about 12 minutes in which the target words were presented in Spanish and English with repetition and on-screen labelling, followed by unrelated free play. In the play-only condition, teachers introduced the target words orally with picture cards and then guided small-group play activities, such as a pretend shop or a picnic, designed to elicit use of the words, without any screen content. In the combined condition, children watched the same episode as the screen-only group and then engaged in the same guided play activities as the play-only group, with teachers explicitly linking play to the episode. Sessions in the combined condition were slightly longer, at approximately 38 minutes, to accommodate both components. Teachers received a two-day training and a session-by-session manual, and fidelity was monitored through fortnightly observation using a checklist; fidelity was acceptable in all classrooms.

3.3 Measures

Receptive vocabulary was assessed with a 40-item picture-selection task in which the child heard an English word and chose among four pictures, drawn from the 72 target words and 8 untaught control words. Expressive vocabulary was assessed with a 30-item picture-naming task in which the child was shown a picture and asked to name it in English. Both measures were administered individually by trained assessors blind to condition, at baseline and within one week of the end of the intervention. Internal consistency was 0.88 for the receptive and 0.86 for the expressive measure at post-test. Home English exposure was measured through a parent questionnaire on the frequency of English media use and English-speaking contacts at home. Spanish vocabulary was assessed at baseline with a standardised measure as a covariate.

3.4 Analysis

Because children were nested in classrooms, two-level multilevel models were estimated with post-test vocabulary as the outcome, condition and city as fixed effects, baseline vocabulary, age, home English exposure and Spanish vocabulary as covariates, and random intercepts for classrooms (Snijders & Bosker, 2012). Condition by city interactions were tested to address H3. Planned contrasts compared the combined condition with each single-component condition and play-only with screen-only. Effect sizes were computed as standardised mean differences on adjusted post-test scores. Attrition was 5.4% and did not differ by condition; missing post-test data were handled by full information maximum likelihood.

4. Results

4.1 Descriptive statistics

Baseline vocabulary did not differ significantly across conditions in either city, though children in Madrid had somewhat higher baseline English vocabulary and home exposure than children in Mexico City, as expected. Table 2 presents baseline and post-test means for receptive and expressive vocabulary by condition and city. All conditions showed gains from baseline to post-test, but the size of the gains varied markedly. The combined condition showed the largest gains in both measures and both cities, play-only showed intermediate

gains with a relatively stronger expressive component, and screen-only showed the smallest gains, concentrated in receptive vocabulary. Control words that had not been taught showed negligible change in any condition, indicating that the gains were specific to the intervention content rather than to general test familiarity.

Table 2: Baseline and post-test vocabulary means (SD) by condition and city

Condition and city	Receptive baseline	Receptive post-test	Expressive baseline	Expressive post-test
Screen only, Mexico City	8.6 (4.0)	15.9 (5.8)	3.1 (2.4)	6.2 (3.5)
Screen only, Madrid	10.3 (4.3)	17.8 (6.0)	4.0 (2.7)	7.4 (3.8)
Play only, Mexico City	8.9 (4.4)	18.7 (6.2)	3.3 (2.5)	10.1 (4.4)
Play only, Madrid	10.5 (4.5)	20.4 (6.3)	4.2 (2.8)	11.3 (4.6)
Combined, Mexico City	8.5 (3.9)	22.1 (6.5)	3.0 (2.3)	13.6 (4.9)
Combined, Madrid	10.0 (4.2)	23.6 (6.4)	3.9 (2.6)	14.8 (5.0)

Source: Authors' study data (2026)

4.2 Hypothesis testing

Table 3 presents the multilevel model results and planned contrasts. For receptive vocabulary, the combined condition produced significantly higher adjusted post-test scores than screen-only (difference 5.9 words, $p < 0.001$, $d = 0.62$) and than play-only (difference 3.2 words, $p = 0.003$, $d = 0.34$). For expressive vocabulary, the combined condition again exceeded screen-only (difference 7.3 words, $p < 0.001$, $d = 0.71$) and play-only (difference 3.4 words, $p = 0.001$, $d = 0.36$). H1 was therefore supported for both outcomes. The contrast between play-only and screen-only was significant for expressive vocabulary (difference 3.9 words, $p < 0.001$, $d = 0.41$) but not for receptive vocabulary (difference 2.7 words, $p = 0.052$, $d = 0.28$), supporting H2. City had a significant main effect on both outcomes, reflecting the higher exposure of Madrid children, but the condition by city interaction was not significant for either outcome, supporting H3: the pattern and approximate magnitude of condition effects were the same in the two cities. Home English exposure was a significant covariate for receptive but not expressive vocabulary.

Table 3: Multilevel model estimates and planned contrasts for post-test vocabulary ($N = 260$, 12 classrooms)

Effect	Receptive: estimate (SE)	p	Expressive: estimate (SE)	p
Baseline vocabulary (covariate)	0.71 (0.06)	< 0.001	0.64 (0.07)	< 0.001
City (Madrid = 1)	1.6 (0.7)	0.028	1.2 (0.5)	0.024
Home English exposure	1.4 (0.6)	0.019	0.7 (0.5)	0.148
Combined vs screen only	5.9 (1.0)	< 0.001	7.3 (0.9)	< 0.001
Combined vs play only	3.2 (1.0)	0.003	3.4 (0.9)	0.001
Play only vs screen only	2.7 (1.0)	0.052	3.9 (0.9)	< 0.001
Condition $\times$ city interaction		0.61		0.44
Classroom variance (ICC)	0.07		0.06	

Source: Authors' study data (2026)

4.3 Observational findings

Fidelity observations recorded the number of target-word productions by children during each observed session. In the play-only condition, children produced a mean of 14.2 target-word tokens per session, and in the combined condition 21.7, while in the screen-only condition, where play was unrelated, productions were rare at 2.1 per session. Observers noted that in the combined condition children frequently referred to characters and events from the episode during play, using the narrative as a shared frame that reduced the teacher's need to prompt. Teachers in the combined condition reported in post-intervention interviews that the episode gave children a common reference point that made it easier to involve quieter children in play. Teachers in the screen-only condition reported that children enjoyed the episodes but rarely used the words spontaneously afterwards.

5. Discussion

The findings extend our earlier research (Rodriguez, 2025) in a specific and practically important direction. That study found that peer interaction predicted second language vocabulary growth more strongly than screen exposure and that screens contributed mainly to receptive knowledge; the present study reproduces that pattern in the comparison of play-only and screen-only conditions, and then demonstrates that structured screen content adds substantial value when it is embedded in guided peer play. The combined condition outperformed both single-component conditions on both receptive and expressive vocabulary, with effects relative to screen-only that would be regarded as moderate to large in educational research. Screens, in other words, are neither a substitute for interaction nor irrelevant to it; their contribution depends on the activity in which they are embedded.

The results are consistent with the joint media engagement framework (Takeuchi & Stevens, 2011) and with evidence that preschoolers learn from video when it is designed with developmental principles and coupled with interaction (Roseberry et al., 2014). They also align with the guided play literature (Weisberg et al., 2016; Toub et al., 2018), and the observational data suggest a mechanism: the screen episode supplied a shared narrative and a set of labelled referents that made subsequent play more focused and more inclusive, increasing the number of opportunities for children to produce the target words. In sociocultural terms (Vygotsky, 1978), the episode functioned as a cultural tool that structured joint activity, and the play provided the scaffolded production that expressive vocabulary requires. The finding that screen-only exposure produced gains mainly in receptive vocabulary while play added expressive gains mirrors the distinction between hearing and using language that runs through the bilingual development literature (Hoff et al., 2012).

The cross-national consistency of the effects is notable. Madrid children entered with more English exposure and higher baseline vocabulary than Mexico City children, and the main effect of city reflects this, but the condition effects were of similar magnitude in both cities. This suggests that the combined approach does not depend on ambient exposure and may be applicable in a wide range of Spanish-speaking contexts, including those where children have little contact with English outside school. The finding that home exposure predicted receptive but not expressive vocabulary further indicates that expressive gains derive from the interactive component of the intervention rather than from background exposure.

For preschool practice, the findings support a purposeful rather than restrictive stance toward screen media. Content should be structured around explicit vocabulary targets, and viewing should be followed immediately by guided play that uses the content as a frame. The additional session time required for the combined condition was modest, at around eight

minutes, and teachers found the approach manageable after brief training. Public preschools in both cities operate with limited resources, and the intervention used freely available episodes and ordinary classroom materials. The results also carry a caution for parents and policymakers who regard screen exposure alone as a route to bilingualism: in the absence of interaction, structured content produced only modest gains, concentrated in recognition rather than use, which is consistent with the conclusions of Madigan et al. (2020) that educational content is beneficial but limited.

6. Conclusion

This cross-national study of 260 preschoolers in Mexico City and Madrid examined how structured bilingual screen content combined with guided peer play affects second language vocabulary growth. The combined condition produced the largest gains in receptive and expressive vocabulary, guided peer play alone outperformed screen content alone on expressive vocabulary, and the pattern of effects was consistent across the two cities despite differences in home exposure to English. The study extends our earlier findings on screen time and peer interaction by showing that the two are complementary when screen content is structured and embedded in interactive play, and it offers a practical model for early childhood educators.

Limitations include the quasi-experimental assignment of a small number of classrooms, which limits the precision of classroom-level variance estimates, the twelve-week duration, and the assessment of vocabulary limited to taught words without a delayed follow-up. The intervention was delivered with research support that regular classrooms may lack. Future research should include delayed assessment to examine retention, test the approach with languages other than English and in rural settings, examine home-based versions in which parents rather than teachers guide play, and investigate which features of screen content most effectively support subsequent play.

Acknowledgment

The authors thank the participating preschools, teachers and families in Mexico City and Madrid, and acknowledge institutional support from UNAM and the Universidad Politécnica de Madrid.

References

- Anderson, D. R., & Subrahmanyam, K. (2017). *Digital screen media and cognitive development*. *Pediatrics*, 140(Suppl. 2), S57–S61. <https://doi.org/10.1542/peds.2016-1758C>
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1–27. <https://doi.org/10.1017/S0305000910000759>
- Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). *Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning*. *Proceedings of the National Academy of Sciences*, 100(15), 9096–9101. <https://doi.org/10.1073/pnas.1532872100>
- Madigan, S., McArthur, B. A., Anhorn, C., Eirich, R., & Christakis, D. A. (2020). Associations between screen use and child language skills: A systematic review and meta-analysis. *JAMA Pediatrics*, 174(7), 665–675. <https://doi.org/10.1001/jamapediatrics.2020.0327>

- Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, 1(3), 213–237. <https://doi.org/10.1075/lab.1.3.01par>
- Rodriguez, S. (2025). Second language acquisition in early childhood: The impact of screen time vs. peer interaction. In *Proceedings of the 2025 International Conference on Creativity, Management, Education, Technology and Sciences* (pp. 251–260). Education Studio (Thailand). <https://doi.org/10.62788/ICCMETS2025-paper25>
- Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, 85(3), 956–970. <https://doi.org/10.1111/cdev.12166>
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.
- Takeuchi, L., & Stevens, R. (2011). *The new coviewing: Designing for learning through joint media engagement*. The Joan Ganz Cooney Center at Sesame Workshop. <https://joanganzcooneycenter.org/>
- Toub, T. S., Hassinger-Das, B., Nesbitt, K. T., Ilgaz, H., Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Nicolopoulou, A., & Dickinson, D. K. (2018). The language of play: Developing preschool vocabulary through play following shared book-reading. *Early Childhood Research Quarterly*, 45, 1–17. <https://doi.org/10.1016/j.ecresq.2018.01.010>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K., & Klahr, D. (2016). Guided play: Principles and practices. *Current Directions in Psychological Science*, 25(3), 177–182. <https://doi.org/10.1177/0963721416645512>
- World Health Organization. (2019). *Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age*. WHO. <https://www.who.int/publications/i/item/9789241550536>

Mock manuscript generated for ICCMETS 2026 portal testing — not a real submission.