

Embodied and Visual Vocabulary Learning with Very Young EFL Learners: A Qualitative Micro-Case of TPR-Supported Instruction

Arvan Permana Habib¹; Fahriany Fahriany^{2*}

^{1,2}Universitas Islam Negeri Syarif Hidayatullah, Indonesia

*Corresponding author, e-mail: fahriany@uinjkt.ac.id

Abstract

Very young learners require vocabulary instruction that makes new words concrete, meaningful, and physically engaging. This qualitative micro-case examines how visual media and Total Physical Response (TPR) mediated vocabulary learning for two Indonesian children: one kindergarten learner and one second-grade elementary learner. Field notes from home-based learning sessions were reanalyzed thematically with attention to meaning construction, embodied response, attention, and learner-initiated interaction. The findings show that pictures and video scenes provided initial semantic hypotheses, while gesture and action allowed learners to test and demonstrate emerging meanings. Misinterpretations, such as identifying a shark's mouth rather than its teeth, became productive moments when visual ambiguity was followed by teacher prompting and physical response. TPR also generated short episodes of playful participation, although attention remained unstable. The study argues that visual input and embodied action are complementary forms of scaffolding rather than independent techniques, and that their value depends on contingent teacher mediation.

Keywords: embodied learning; total physical response; visual media; vocabulary learning; young EFL learners.

How to Cite: Habib, A. P. & Fahriany, F. (2025). Embodied and Visual Vocabulary Learning with Very Young EFL Learners: A Qualitative Micro-Case of TPR-Supported Instruction. *Charta Educa: Jurnal Kajian Pendidikan*, 1(3). 98-106.



This is an open access article distributed under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2025 by author.

Introduction

Vocabulary is one of the earliest and most visible challenges faced by children who begin learning English as a foreign language. A young learner may recognize a picture, imitate a sound, repeat a word, or respond to a classroom command long before being able to explain grammatical rules. For this reason, vocabulary instruction in early English learning is not simply a matter of memorizing lists. It involves helping children connect unfamiliar linguistic forms to objects, actions, categories, routines, and experiences that already make sense in their world. The pedagogical problem is therefore one of meaning-making: how can an English word become sufficiently concrete for a child to notice, interpret, remember, and use?

The answer is complicated by the developmental characteristics of young learners. Children differ from adolescent and adult language learners in attention span, literacy development, metalinguistic awareness, emotional sensitivity, and dependence on contextual support. Nunan (2016) emphasizes that teaching young learners requires an understanding of age-related learning characteristics rather than the simple transfer of methods developed for adults. Studies of English teaching to young learners similarly show that teachers need flexible strategies because children may be enthusiastic and imitative while also becoming distracted quickly, especially when language is presented abstractly or through lengthy verbal explanation (Oktavia et al., 2022). These characteristics are particularly important in very small-scale or home-based learning, where the boundaries between lesson, play, conversation, and everyday activity are porous.

Visual media are frequently recommended because they reduce the distance between an unfamiliar word and an interpretable referent. Pictures, objects, video, and other multimodal resources can provide contextual cues that enable learners to infer meaning before receiving a translation. Sofian and Anggraeni (2021) found that visual media can support English teaching to young learners by attracting attention and making classroom input more concrete. More recent Indonesian research has continued to show the usefulness of image-based and audio-visual materials for vocabulary instruction. Brilianti and Sugirin (2024), for example, reported improvement in young learners' vocabulary scores after the use of puppet-picture storytelling, while Setiawan and Firdaus (2024) argue that audio-visual materials can make vocabulary more engaging and memorable by combining visual and auditory cues. Rahman, Hassan, and Manu (2024) likewise highlight picture-based English materials as a resource for vocabulary development in elementary settings.

Visual cues, however, do not guarantee correct interpretation. A picture can contain many possible referents. A child shown a shark while the target word is teeth may attend first to the animal, the mouth, the color, or the dramatic shape of the jaw. The image therefore offers a semantic field rather than a single meaning. This point is important because visual instruction is sometimes discussed as though the picture automatically carries the meaning of the word. In practice, learners select, test, and revise interpretations. The teacher's questions, gestures, timing, repetition, and responses to confusion can determine whether visual information becomes usable linguistic knowledge.

Total Physical Response (TPR) provides a second route to meaning by linking spoken language to bodily action. Asher's original formulation of TPR treated comprehension through physical response as an important pathway in early language learning. Rather than requiring immediate verbal production, learners listen to commands and demonstrate understanding by acting. The approach has remained attractive for young learners because movement can turn language into an event rather than an abstract explanation. Xie (2021) found that TPR can be effective in English teaching to young learners while also noting implementation challenges and the importance of combining it with other participatory activities. Indonesian studies have similarly reported that action-based instruction can support vocabulary learning. Setiawan, Nurhajati, and Sulistyani (2022) describe TPR as a way to teach vocabulary through embodied participation, and Riyanti and Rustipa (2024) report active learner responses when TPR was used in kindergarten vocabulary teaching.

The 2024 literature shows continuing interest in TPR but also a movement toward more nuanced combinations of action, personalization, and media. Liu, Chen, and Chen (2024) compared traditional and personalized TPR strategies for young EFL learners and demonstrated that the design of the physical response matters rather than TPR functioning as a single uniform treatment. Jazila and Megawati (2024), working with kindergarten learners, found that children responded positively to TPR-based vocabulary activities and emphasized observation of classroom participation and comprehension. Sainyakit (2024), in a literature study, synthesized evidence indicating that TPR can support vocabulary development across several contexts. Ninsiana et al. (2024) also used TPR in an English vocabulary community-service setting, illustrating its practical appeal in contexts where young learners need highly accessible forms of participation.

These studies are useful, but much of the literature evaluates TPR or visual media in terms of outcomes such as vocabulary gains, interest, or participation. Less attention is often given to the micro-process through which a child moves from seeing and hearing a word to constructing a tentative meaning, asking for clarification, acting, and linking the word to personal experience. In a small qualitative study, the value lies precisely in examining these episodes closely. Such analysis cannot establish the effectiveness of an intervention for a population, but it can illuminate how instructional resources function in situated interaction.

The study reported in this article involved only two learners, a kindergarten pupil and a second-grade elementary pupil, in a home-based setting. It used field notes to document sessions in which visual materials, including English-learning videos, were combined with TPR-style instructions. The reported study described the design as narrative inquiry and concluded that visual media and TPR helped vocabulary learning. In this analysis, the empirical scope is stated more conservatively. The study is treated as a qualitative micro-case of instructional interaction. The purpose is not to claim causal vocabulary improvement from two cases, but to examine what the observed episodes reveal about the complementary roles of visual cues, embodied response, teacher mediation, and learner attention.

This reframing is also consistent with contemporary concerns about multimodal vocabulary instruction. Young learners increasingly encounter English through video, digital stories, images, games, and short-form media. The pedagogical issue is not whether technology is inherently motivating, but how multimodal information is coordinated with teacher guidance. Visual media can increase available cues, yet too many cues may compete for attention. Physical response can embody meaning, yet movement can also turn into off-task play if the task lacks structure. An analytically useful question therefore concerns the relationship among these elements rather than their isolated effects.

The study addresses two questions. First, how did the two learners use visual cues and TPR-supported action to construct or demonstrate meanings of unfamiliar English vocabulary? Second, what interactional conditions appeared to support or interrupt that process? By focusing on episodes of inference, confusion, physical demonstration, questioning, and playful attention, the article contributes a process-oriented account of early vocabulary learning. The central argument is that visual media and TPR are most productive when understood as complementary forms of scaffolding: visuals open possibilities for semantic inference, action makes comprehension observable, and teacher mediation helps learners select, test, and stabilize meanings. The qualitative design follows established guidance on bounded educational inquiry and context-sensitive interpretation (Creswell & Creswell, 2023; Wa-Mbaleka & Rosario, 2022).

Method

This study uses a qualitative micro-case design based on the original field-note study. The earlier publication described the approach as narrative inquiry because it documented the researcher's teaching experience with two very young learners. That orientation is retained in the sense that lived instructional episodes are the principal data source, but the revised methodological description avoids implying a broader narrative-life-history analysis. The analytic unit is the learning episode: a short sequence in which a vocabulary item was presented, interpreted, questioned, acted upon, or connected with an object or experience.

The participants were two children in different stages of early schooling. One was enrolled in kindergarten and the other in the second grade of elementary school. The learning sessions took place in the learners' residence rather than in a formal school classroom. This setting is important for interpretation because the interaction was intimate and flexible, and distractions from objects, play, and conversation were part of the environment. The study therefore does not represent a conventional whole-class TPR lesson. It provides a close view of two learners interacting with an instructor and multimodal materials in a home-based educational setting.

Data consisted of field notes produced during or immediately around the learning sessions. The notes recorded the use of English-learning videos and visual images, learners' guesses about word meanings, their questions, their physical responses to TPR-style prompts, and observable changes in attention. The study provides concrete episodes, including work with the words tooth/teeth and with color vocabulary. Because no audio transcripts, standardized vocabulary tests, or longitudinal retention measures were reported, the analysis does not infer gains beyond what the field notes support. Statements such as "the learner understood" are used only when an observed response provided evidence of comprehension in that moment, for example by pointing to, showing, or acting on the correct referent.

The material was reanalyzed thematically. First, all source descriptions were reviewed to identify episodes in which visual information, verbal explanation, physical response, learner questions, and off-task behavior occurred. Second, these episodes were coded according to their pedagogical function: visual inference, ambiguity or misinterpretation, clarification, embodied demonstration, personal connection, repetition, and attention shift. Third, codes were grouped into four themes presented in the Results section: visual media as semantic hypothesis space, embodied response as a comprehension display, contingent teacher mediation, and the instability of attention in playful learning. The Discussion relates these themes to recent studies of TPR and multimodal vocabulary instruction.

Credibility is supported through analytic restraint and transparency about the data's limitations. The article preserves the observable sequence of key episodes rather than reconstructing unrecorded dialogue. It distinguishes participant behavior from interpretation and avoids causal language that would require experimental comparison. The study also treats the two learners as cases rather than a sample intended for statistical generalization. Its contribution is analytic: identifying processes that can inform hypotheses and instructional design in larger studies. Because the study report does not provide detailed information about formal consent procedures, session duration, number of sessions, or participant demographic characteristics beyond school level, these details are not invented in the revised account and should be verified from the original field records before any new submission.

Results and Discussion

Visual cues created semantic hypotheses rather than automatic meanings

The clearest pattern in the field notes was that the learners used images and video scenes as evidence for guessing what an unfamiliar English word might mean. This was visible when a target word appeared while a picture or moving image was shown. The children did not wait passively for translation in every instance. They scanned the visual scene, selected a salient object, and proposed an interpretation. Such behavior suggests that

visual media functioned as a semantic hypothesis space: the image narrowed the possibilities but did not uniquely determine meaning.

The episode involving tooth/teeth illustrates this process. A shark was shown with its mouth open. The visual contained several plausible referents, including the shark itself, the mouth, the jaw, and the teeth. The learners initially associated the target with *mulut* (mouth). That answer was understandable because the mouth was a visually dominant feature of the picture. The response therefore should not be treated simply as an error. It revealed where attention had been allocated and what relationship the learners perceived between the English form and the visual field.

Once the mismatch became apparent, the instructional interaction changed. Rather than abandoning the visual, the teacher used it as a starting point for more focused prompting. The target referent was identified, and the learners were invited to relate the shark's teeth to their own. The sequence moved from image-based guessing to embodied and personal reference. This movement is pedagogically important because it shows why visuals are not self-explanatory. Their instructional value depends on the questions and actions that help learners isolate the relevant feature.

Similar processes occurred when learners encountered other vocabulary items in video materials. The field notes indicate that they sometimes asked directly for meaning when the visual context was insufficient. This learner-initiated questioning is significant. It suggests that confusion did not always result in withdrawal. At some points, uncertainty generated a request for clarification. In a small learning setting, such questions gave the instructor immediate information about which cues were working and which were not.

The pattern is consistent with picture-based vocabulary research, which emphasizes the value of concrete visual support while also implying a role for teacher mediation. Rahman et al. (2024) report that picture-based materials can strengthen vocabulary mastery by making lexical meanings more accessible. Brilianti and Sugirin (2024) similarly show that puppet pictures embedded in storytelling can support vocabulary acquisition. The cases add a process-level observation: the benefit does not arise because a picture contains a fixed answer. Learners still need to determine which aspect of the picture is linguistically relevant.

Physical response made emerging comprehension observable

A second pattern was that TPR-style prompts enabled the instructor to see whether a learner could connect the spoken word with an appropriate action or referent. After the tooth/teeth meaning was clarified, the teacher asked the learners whether they had teeth and invited them to show them. One learner opened his mouth and displayed his teeth. The learners also began talking about having lost some teeth. The target vocabulary had moved from a two-dimensional shark image to the children's own bodies and experiences.

This transition illustrates an important function of TPR in very early vocabulary learning. A physical response can operate as an informal comprehension check before accurate verbal production is expected. The child does not need to define the word in English. Showing the relevant body part provides evidence that the sound form has been linked, at least temporarily, to an appropriate meaning. In instructional terms, action externalizes comprehension.

The same principle appeared in the color activity used near the end of a session. The learners were shown objects of similar shape but different colors and were asked about the English color terms. Their initial knowledge of English color vocabulary was limited. They were then asked to find or show possessions and nearby objects that had the same color. This transformed an abstract naming exercise into a search-and-show task. The learners reportedly became more enthusiastic when they could move, look around the room, and bring their environment into the activity.

The observation aligns with recent TPR studies. Jazila and Megawati (2024) report active responses from kindergarten learners when vocabulary was connected with physical action. Riyanti and Rustipa (2024) likewise describe TPR as encouraging participation in kindergarten vocabulary learning. Liu et al. (2024) extend the argument by showing that the way TPR is personalized can influence efficiency. The present micro-case helps explain why personalization may matter: an action tied to the learner's own body or possessions gives the word a personally available referent.

Physical activity did not eliminate the need for language. Rather, it created a sequence in which listening, action, and short verbal exchanges reinforced one another. The teacher's spoken instruction established the target expression, the learner's response displayed an interpretation, and subsequent conversation provided additional contextualization. TPR therefore functioned less as a stand-alone method than as one component of an interactional cycle.

Teacher mediation connected seeing, hearing, and doing

A third finding concerns the instructor's mediating role. The learning episodes were not simply instances of children watching videos or responding to commands. Progress occurred when the instructor noticed the learners' interpretation and adjusted the next move. In the shark episode, an incorrect guess prompted a

narrowing of attention. After clarification, a body-based question was used to confirm meaning. In the color activity, the task was extended from naming to locating similar colors in the immediate environment.

This contingent response is important because the same media could produce different learning opportunities depending on teacher action. A video can be played without interaction, leaving learners to process an uncontrolled stream of language. It can also be paused, questioned, repeated, and connected to an object that the learner can touch. The field notes suggest that the latter form of mediation was more likely to elicit learner talk and physical response.

Visual and physical resources were thus complementary. The visual established a context; the teacher directed attention; the action tested or deepened the interpretation. This relationship can be described as multimodal scaffolding. Each mode contributed different information, but no single mode guaranteed comprehension. The visual might be ambiguous, the command might be misunderstood, and the action might be imitative rather than semantically grounded. The teacher's role was to coordinate the modes and interpret the learner's response.

Recent work on audio-visual vocabulary materials makes a similar point in design terms. Setiawan and Firdaus (2024) emphasize that multimedia resources combine video, animation, and interactive features to support retention and comprehension. Yet an educational resource becomes instruction only through tasks that direct learners toward particular meanings. Brilianti and Sugirin (2024) integrated pictures within storytelling rather than displaying pictures in isolation. Ninsiana et al. (2024) used TPR in guided vocabulary training. Across these approaches, learning depends on the relationship between resource and participation.

The micro-case also showed that learners sometimes extended the interaction beyond the target item. After understanding a word, they asked questions or offered personal information. Such expansions may appear tangential, but in very young learner instruction they can indicate that the new word has become connected to an existing conceptual network. Talking about missing teeth after learning tooth/teeth is not merely off-topic; it is a sign that the word has activated autobiographical knowledge. Teacher judgment is required to decide when such talk enriches learning and when it pulls attention away from the lesson.

Attention was productive but unstable

The fourth theme was the instability of attention. The learners did not remain continuously focused. The original field notes describe the learning process as taking longer because the children played with one another, turned away, and engaged in their own activities. This is not a minor methodological detail. It is central to understanding the observed instructional process.

Visual media and movement appeared to generate enthusiasm at particular moments, especially when learners could show objects or act on a command. However, engagement was episodic rather than constant. A stimulating task could draw attention back, but the same freedom of movement that made TPR enjoyable could also create opportunities for distraction. The home environment likely amplified this tension because familiar objects and play possibilities remained immediately available.

This finding cautions against equating visible activity with sustained learning. TPR is often attractive because children move and respond, but movement can have different functions. It may represent comprehension, imitation, excitement, or unrelated play. The teacher needs to observe whether the action remains tied to the linguistic target. Similarly, video can hold attention through animation or novelty without ensuring that learners notice the intended word.

Young learner research has long emphasized short attention spans and the need for varied activities. More recent work reinforces the importance of engaging but purposeful media. Izati (2024) discusses fun learning activities as a way to increase interest in elementary English learning, while Bakhtiyarova (2024) highlights the appeal of TPR for active language instruction. The cases suggest that variation should be organized around a learning sequence rather than constant novelty. Short cycles of visual input, prediction, action, feedback, and reuse may be more productive than extended exposure to a single medium.

The embodied component is consistent with the action-based logic of Total Physical Response (Asher, 1977), while recent work with very young learners emphasizes the importance of age-appropriate, multimodal, and interaction-rich language experiences (Ghaffari Saravi et al., 2023). Research on video-supported English learning likewise shows that visual media can extend exposure and engagement when they are integrated with purposeful tasks rather than used passively (Yaacob et al., 2021).

Discussion

Visual meaning-making is inferential, not automatic

The first contribution of this micro-case is to complicate a simple claim frequently made about visual media: that pictures make vocabulary meaning clear. The observations show that images can make meaning available without making it determinate. The shark picture narrowed the semantic field but initially led learners to the

mouth rather than the teeth. This is exactly the kind of moment that outcome-oriented studies may not capture, yet it reveals a core process in vocabulary learning: learners infer, compare, and revise.

This interpretation helps explain why visual materials are especially useful when paired with interaction. Rahman et al. (2024) found positive value in picture-based materials, and Brilianti and Sugirin (2024) showed that pictures embedded in a narrative context can support vocabulary acquisition. In both cases, the pedagogical structure around the image matters. The present findings suggest that teachers should not ask only whether a visual is attractive or relevant; they should ask whether the intended referent is perceptually salient and whether prompts are available when several interpretations are possible.

For classroom practice, one implication is to design visual prompts that progressively focus attention. A teacher can begin with an open question such as "What can you see?", then narrow the target through pointing, contrast, or action. For teeth, contrasting mouth/teeth, asking learners to touch their teeth, or showing one tooth and several teeth can make both semantic and morphological distinctions visible. Such sequences turn a mistaken guess into diagnostic information.

Embodiment works as assessment as well as instruction

A second contribution is the recognition that TPR provides a form of immediate formative assessment. When a child performs an appropriate action, the teacher gains evidence about comprehension without requiring a translation or written answer. This feature is especially valuable for very young learners whose productive English and literacy may lag behind conceptual understanding.

Jazila and Megawati (2024) describe observed comprehension and participation in TPR-based learning, while Liu et al. (2024) demonstrate that TPR can be adapted and personalized. The cases show a mechanism connecting these findings: action transforms an internal interpretation into a visible performance. The teacher can then respond. If a learner points to the wrong object, the misconception becomes observable; if the learner performs the action correctly but cannot yet say the word, the teacher can distinguish receptive from productive knowledge.

This distinction is pedagogically important. Vocabulary knowledge is not binary. Learners may recognize a word before producing it, identify a referent before spelling it, or respond to a command before using the word spontaneously. TPR fits this developmental progression because it allows comprehension to precede oral production. However, TPR should not become an endpoint. Repeated cycles can gradually move from teacher command to learner command, from showing to naming, and from naming to using the word in a short utterance.

Multimodal scaffolding depends on contingent mediation

The strongest theoretical implication of the study is that visual media and TPR should be understood relationally. The question is not whether visuals or TPR are "effective" in isolation. The useful unit is the mediated sequence: image or video provides a cue, the learner interprets it, the teacher responds to that interpretation, action makes meaning visible, and language is recycled in a personally meaningful context.

This relational view is consistent with the increasing use of multimodal resources in young learner vocabulary teaching. Setiawan and Firdaus (2024) propose audio-visual materials that combine several channels; Sainyakit (2024) synthesizes evidence for TPR; Brilianti and Sugirin (2024) use puppet pictures and storytelling. The variety of successful arrangements suggests that no single medium is sufficient. What matters is how the modes distribute the cognitive work of meaning-making.

A practical risk is sensory or attentional overload. A video may contain music, movement, multiple objects, subtitles, and rapid speech. For adults, such richness may seem engaging. For a young learner, it can obscure the target. Teachers can reduce this risk by pausing, replaying, freezing frames, pointing, or simplifying the scene. Physical response can then help anchor the target meaning in action.

The data also reinforce the importance of contingent teaching. The instructor did not follow an entirely fixed script. Questions emerged from the learners' responses. This is characteristic of high-quality early learning: instruction is planned, but the route through the plan depends on what children notice and misunderstand. A two-learner setting makes this responsiveness relatively easy. Larger classrooms require different organizational strategies, such as pair demonstration, whole-class gestures, small-group picture tasks, and rapid checks that allow the teacher to sample comprehension.

Engagement should be interpreted cautiously

The study also provides a useful caution about motivation and engagement. The learners were described as enthusiastic in some tasks, particularly when movement and personal objects were involved, but they also played, turned away, and lengthened the session. It would therefore be misleading to conclude that TPR or video created uniformly high engagement. The more defensible claim is that they generated episodes of re-engagement and meaningful participation.

This distinction matters in educational research. Visible excitement is often treated as evidence that a method works. Yet children can enjoy an activity without learning the intended vocabulary, and they can learn

during quieter moments that do not look highly animated. Researchers and teachers should distinguish behavioral engagement, linguistic participation, and evidence of comprehension. A child moving energetically is not necessarily responding semantically; a child quietly pointing to the correct image may provide stronger evidence of learning.

The implication for lesson design is to use movement strategically. TPR works best when the action has a direct semantic relationship with the target, when instructions are brief, and when the sequence includes opportunities to check meaning. Short activity cycles can respect children's attention without allowing the lesson to become a succession of entertainment events.

Implications for early English pedagogy

Several practical principles emerge from the findings. First, teachers should select visuals in which the target referent is identifiable but still allow meaningful inference. Second, they should expect ambiguity and prepare prompts rather than treating an incorrect visual guess as failure. Third, physical response can be used as a low-language-demand comprehension check. Fourth, personal objects and body-based examples can help move vocabulary from screen to experience. Fifth, instruction should recycle the word across modes: hear it, see it, act on it, point to it, say it, and encounter it again.

These principles are compatible with the recent Indonesian literature. Ninsiana et al. (2024) demonstrate the accessibility of TPR in community-based teaching; Riyanti and Rustipa (2024) show its use with kindergarten learners; Jazila and Megawati (2024) document positive learner perceptions; and Rahman et al. (2024) support the use of pictures in elementary vocabulary instruction. The value of the present study lies in showing how these techniques can interact in real-time teacher-learner exchanges.

The analysis also supports cautious use of digital media. English-learning videos were helpful as sources of images and language, but the platform itself was not the pedagogical intervention. The teacher's pausing, questioning, explanation, and extension of video content transformed media exposure into a lesson. This distinction becomes increasingly important as children have abundant access to English media outside school. Exposure can create opportunities, but guided attention determines which opportunities become learning.

Limitations and directions for further research

The limitations of the study are substantial and should define how its findings are used. There were only two participants, the setting was home-based, and the data were field notes rather than audio- or video-recorded transcripts. The study report does not provide standardized pretest-posttest measures, delayed retention tests, session frequency, or detailed participant profiles. The study therefore cannot establish that TPR and visual media caused vocabulary gains, nor can it compare their independent contributions.

The small scale is nonetheless useful for generating more precise questions. Future studies could video-record interaction to analyze gaze, gesture, pointing, teacher prompts, and learner repair in detail. Researchers could compare visual-only, action-only, and integrated sequences while retaining qualitative observation of how meaning is negotiated. Delayed tests could examine whether words connected with personal action are remembered more reliably than words encountered only in video. Larger classroom studies could investigate how teachers manage TPR when twenty or thirty children move simultaneously and how they distinguish imitation from comprehension.

Research could also explore learner differences. Some children may respond strongly to physical action, while others may prefer visual observation or verbal repetition. Personalized TPR, as examined by Liu et al. (2024), offers one direction. Another is to investigate how age, first-language literacy, attention regulation, and familiarity with digital media shape the usefulness of multimodal vocabulary tasks. Methodologically, future work should preserve process data alongside outcome measures. A higher posttest score tells us whether performance changed; an interactional record helps explain how. Combining the two can produce a richer account of why a technique works for some words, learners, or moments but not others.

Conclusion

The two cases examined in this study show that visual media and TPR can create complementary pathways for very young learners to make sense of unfamiliar English vocabulary. Visuals supported initial inference, but they did not automatically determine the intended meaning. Physical response made emerging comprehension observable and allowed new words to be connected with the learners' own bodies and surroundings. The most productive episodes occurred when teacher mediation linked seeing, hearing, questioning, and doing.

The study also shows why claims about effectiveness must remain proportionate to the evidence. Two learners and field-note data cannot support broad conclusions about vocabulary achievement. What the cases provide is a close view of instructional process. Misinterpretation, clarification, embodied response, personal connection, and temporary distraction were all part of learning rather than noise surrounding it.

For practice, the implication is to treat visual media and TPR not as competing techniques or motivational add-ons, but as coordinated forms of scaffolding. Teachers can use visuals to invite semantic hypotheses, action to test comprehension, and responsive prompts to refine meaning. When vocabulary instruction follows this cycle, multimodal resources become more than attractive materials: they become tools for making language interpretable, observable, and personally meaningful.

Declarations

Author contribution

Author 1: developed the main research idea and objectives, designed the research methodology, conducted data collection and investigation, performed the data analysis, and prepared the first draft of the manuscript. Author 2: contributed to the development and implementation of the research methodology, supported the data collection and investigation process, and reviewed and revised the manuscript to improve its academic quality, clarity, and structure, provided input into the research concept and direction, supervised and guided the overall research process, reviewed and improved the manuscript, and contributed to obtaining financial support or research funding

Funding

The research received no specific grant from any funding agency.

Conflict of Interest

The authors declare that there is no conflict of interest related to the publication of this article.

Use of generative AI

Generative AI were used solely for translation and language editing of the manuscript. The authors take full responsibility for the content and accuracy of the article.

Acknowledgments

Authors would like to express their sincere gratitude to all parties who contributed to and supported the completion of this study. Their valuable assistance, support, and encouragement are greatly appreciated.

References

- Asher, J. J. (1977). *Learning another language through actions: The complete teacher's guidebook*. Sky Oaks Productions.
- Bakhtiyarovna, M. U. (2024). Teaching English language using Total Physical Response method. *Miasto Przyszłości*, 46, 825-828.
- Brilianti, A. E., & Sugirin, S. (2024). Boosting young EFL learners' English vocabulary acquisition by utilizing puppet pictures as storytelling media. *Journal of Foreign Language Teaching and Learning*, 9(2), 161-179. <https://doi.org/10.18196/ftl.v9i2.21622>
- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE.
- Ghaffari Saravi, Z., Rashtchi, M., & Keyvanfar, A. (2023). Teaching English to very young learners: A case study of two three-year-old twins. *Language Related Research*, 14(5), 83-108.
- Izati, L. E. A. (2024). The use of fun learning activity to increasing students' interest in learning English at SD YBPK Semampir. *ELTT*, 10(1), 308-312.
- Jazila, I. N., & Megawati, F. (2024). Using the Total Physical Response method to improve young learners' perceptions of vocabulary mastery. *Pubmedia Jurnal Pendidikan Bahasa Inggris*, 1(4), 13. <https://doi.org/10.47134/jpbi.v1i4.692>
- Liu, P.-L., Chen, C.-J., & Chen, H.-C. (2024). Teaching young EFL learners new vocabulary: A comparison of the efficiency of traditional and personalized TPR strategies. *SAGE Open*, 14(4), 1-14. <https://doi.org/10.1177/21582440241288924>
- Ninsiana, W., Rosidah, R., Siregar, S., Deiniatur, M., Suprihatin, Y., Septiyana, L., & Huda, M. (2024). Introducing the English vocabulary and its pronunciation to young learners through community service training. *Dedikasi: Jurnal Pengabdian Masyarakat*, 6(1), 21-30.
- Nunan, D. (2016). Teaching English to young learners. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 3, pp. 68-81). Routledge.

- Oktavia, D., Mukminin, A., Marzulina, L., Harto, K., Erlina, D., & Holandyah, M. (2022). Challenges and strategies used by English teachers in teaching English language skills to young learners. *Theory and Practice in Language Studies*, 12(2), 382-387.
- Rahman, A., Hassan, A. J., & Manu, S. M. T. (2024). The use of picture-based English materials to improve students' vocabulary mastery for state elementary schools. *Journal of Language and Literature Studies*, 4(1), 135-149.
- Riyanti, A., & Rustipa, K. (2024). The implementation of Total Physical Response to encourage students at Margosari Kindergarten to learn vocabulary. *Jurnal JOEPALLT*, 12(1), 32-44.
- Sainyakit, P. (2024). The students' English vocabulary by using Total Physical Response: A literature study. *International Journal of Education, Language, Literature, Arts, Culture, and Social Humanities*, 2(3), 44-53. <https://doi.org/10.59024/ijellacush.v2i3.918>
- Setiawan, D., Nurhajati, D., & Sulistyani, S. (2022). Teaching vocabulary through Total Physical Response to elementary students. *Journal of Development Research*, 6(2), 158-163.
- Setiawan, M. D., & Firdaus, R. A. (2024). Developing audio-visual materials to teach vocabulary to young learners. *Khalifah: Jurnal Pendidikan Nusantara*, 2(2). <https://doi.org/10.62523/khalifah.v2i2.63>
- Sofian, R. F., & Anggraeni, A. (2021). An analysis of the use of visual media in teaching English to young learners. *PROJECT (Professional Journal of English Education)*, 4(4), 622-627.
- Wa-Mbaleka, S., & Rosario, A. (Eds.). (2022). *The SAGE handbook of qualitative research in the Asian context*. SAGE.
- Xie, R. (2021). The effectiveness of Total Physical Response (TPR) on teaching English to young learners. *Journal of Language Teaching and Research*, 12(2), 293-303.
- Yaacob, A., Amir, A. S. A., Asraf, R. M., Yaakob, M. F. M., & Zain, F. M. (2021). Impact of YouTube and video podcast on listening comprehension among young learners. *International Journal of Interactive Mobile Technologies*, 15(20), 4-19.