The American Journal of Social Science and Education Innovations
ISSN 2689-100X Volume 08 - 2026

Multisensory Anchors in Learning English: How Sound, Sight, and Movement Work Together in Memory and Speech

Kostiuk Tetiana
Linguist, specialist in intercultural communication and language systems,
expert in accelerated foreign language acquisition

Received: 13 Jan 2026 | Received Revised Version: 29 Jan 2026 | Accepted: 15 Feb 2026 | Published: 27 March 2026

Volume 08 Issue 03 2026 | Crossref DOI: 10.37547/tajssei/Volume08Issue03-08

Abstract

Teachers argue endlessly: should students mostly listen, stare at pictures, or physically move when learning English vocabulary? Reality is most classrooms just pick whatever and run with it. Some drill pronunciation for hours. Others spam flashcards. A few remember gestures exist and throw them in occasionally. Nobody thinks much about why. Thing is, nobody's brain evolved processing one sense at a time. Your ancestors who survived? They saw the predator AND heard it AND felt ground vibrations AND started running - all simultaneously. That's baseline human cognition, not some special skill. So the question: does throwing audio, visuals, body movement together - academics label this multisensory scaffolding - actually make vocabulary stick longer and speaking work better than normal one-channel teaching? We dug through experiments from 2008-2024 where researchers tested this on people learning second languages. Hunted specifically for studies tracking how long vocabulary lasts in memory and how well people can produce those words later. What showed up: multisensory teaching beats single-channel pretty reliably across ages. Effect sizes run moderate to large depending how you implement it. Vocabulary learned with gestures+audio+visuals sticks around 3-8 months longer than just hearing words. Speaking accuracy goes up when learners physically act out meanings during learning. Why this works: seems to involve deeper memory encoding through distributed brain networks instead of just repetition effects. What matters critically: gestures have to actually represent meanings (iconicity), timing across senses needs coordination (synchrony), learners doing movements themselves beats watching someone else. What this means practically: pair new words with meaningful gestures and visual supports as standard procedure, not special occasion enrichment. VR might provide structured multisensory environments but jury's still out. Gaps in research: most studies test concrete vocabulary (nouns for objects, action verbs) rather than abstract grammar, and nobody's nailed down optimal dosage or timing yet.

Keywords: multisensory learning, embodied cognition, second language acquisition, gesture-based teaching, vocabulary retention

Cite This Article: Tetiana, K. (2026). Multisensory Anchors in Learning English: How Sound, Sight, and Movement Work Together in Memory and Speech. The American Journal of Social Science and Education Innovations, 8(03), 71–79. https://doi.org/10.37547/tajssei/Volume08Issue03-08

Introduction

Walk into any English classroom and you'll see the same three things over and over. First thing: teacher says a word, students repeat it back like parrots, and everyone acts like getting the pronunciation perfect is the whole point. Second thing: teacher fills the whiteboard with vocabulary words and pictures, students copy everything down in their notebooks, and somehow this is supposed to help "visual learners." Third thing: once in a while they

do some kind of vocabulary game or charades, but it's treated like a fun break, not actual teaching.

So where do these methods even come from? Mostly two places. Teachers teach the way they were taught - if something worked for them as a student, they figure it'll work for their students too. Then they add whatever their training program told them to do. Problem is, training programs don't agree on anything. One program says listening is everything - students learn languages through their ears, period. Another program is all about visual stuff - flashcards, pictures, charts. Some programs mention movement-based learning, but then they brush it off as just a classroom management trick, not a real way people actually learn.

Splitting students into sensory categories seemed logical initially. Humans obviously lean toward particular input channels-some people absorb spoken information rapidly, others need diagrams, still others fidget constantly. Classroom variety prevents mass unconsciousness during lessons. Educational psychology invested enormous effort categorizing learners: you're auditory, you're visual, you're kinesthetic. Diagnostic tools emerged. Teaching guides proliferated. Teachers received training in identifying student types, then customizing instruction accordingly. The system felt scientific, organized, implementable.

Except learning styles theory is bullshit. Researchers hunted for decades trying to prove matching teaching to learning styles actually helps - found nothing (Fallace, 2023). Students like certain inputs better than others, fine. Liking something doesn't equal learning it better though. Person who loves flashcards won't necessarily learn more from pure visual teaching versus mixed approaches. Entire sorting system originated with education theorists in 1900s taking educated guesses that somehow calcified into accepted truth despite lacking any supporting data.

Cognitive neuroscience meanwhile discovered totally different reality. Human thinking is multisensory at core. Brain doesn't have separate audio processor, visual processor, movement processor running independently. Information from senses integrates across multiple levels. Hearing word while seeing referent while making related gesture - inputs merge in distributed neural networks creating richer longer-lasting memory than single input (Shams & Seitz, 2008). Not about student preferences. About memory encoding mechanics.

Embodied cognition research dug deeper into language and brain. Understanding "kick" partially activates motor cortex - same region firing during actual kicking. Abstract concepts work similarly. "Grasping idea" fires some identical neurons as physically grasping objects (Pelkey, 2023). Language isn't abstract symbols floating separate from body. It's embedded in sensorimotor systems.

Not guesswork anymore. Last 15 years saw multiple research teams test whether multisensory teaching actually improves second language learning. They tracked how long vocabulary stuck around in memory, measured speaking accuracy, scanned brains during learning and retrieval. Enough evidence accumulated to ask: does throwing audio, visual, kinesthetic input together genuinely help English learning, or is this just another trendy education fad without real substance?

This review goes through that evidence. Three things we wanted answers for: Does multisensory learning make vocabulary last longer in memory compared to standard audio-only or visual-only teaching? Does it improve speaking - how fast you retrieve words, how accurately you produce them, how fluently you speak? What practical details matter - which combinations of senses work, how timing should go, does making movements yourself beat watching someone else? Point is giving teachers research-backed guidance on whether and how to actually use multisensory scaffolding when teaching English.

Materials and Methods

We searched academic databases (PubMed, PsycINFO, ERIC, Web of Science) for empirical studies published 2008-2024 examining multisensory approaches in second language learning. Search terms combined ("multisensory" OR "multimodal" OR "embodied" OR "gesture-based" OR "kinesthetic") with ("second language" OR "L2 learning" OR "foreign language" OR "English learning") and ("vocabulary" OR "retention" OR "memory" OR "speaking" OR "production"). This yielded 287 potentially relevant articles.

Inclusion criteria: Studies had to (1) test second language learners, not native speakers learning their first language; (2) include at least one condition combining multiple sensory modalities (audio+visual, audio+gesture, audio+visual+gesture); (3) include comparison condition using fewer modalities; (4) measure learning outcomes

through objective tests (vocabulary recall, recognition, production accuracy) rather than just self-reported satisfaction; (5) provide enough methodological detail to evaluate quality. We excluded studies focused solely on first language acquisition, studies without comparison groups, purely theoretical papers, and studies where "multisensory" just meant providing subtitles on videos (which is standard practice, not experimental manipulation).

This left 43 studies forming the evidence base. Most were controlled experiments with random assignment to learning conditions. Sample sizes ranged from 24 to 156 participants. Age groups included children (8-12 years), adolescents, university students, and older adults. Language backgrounds varied but most studied English as second language. Learning materials focused heavily on concrete vocabulary (nouns for objects, verbs for actions) with fewer studies examining abstract vocabulary or grammatical structures.

We extracted data on: learning conditions tested, sensory modalities involved, specific implementation (type of gestures, visual materials, etc.), participant characteristics, testing timeline (immediate versus delayed), outcome measures, and effect sizes where reported. For studies not reporting standardized effect sizes, we calculated these from reported statistics when possible.

We checked if the studies were any good: did they actually randomize people, how many participants showed up, how many quit halfway through, did the tests measure what they were supposed to measure, did they account for obvious stuff like whether someone already knew some vocabulary or was just more motivated. Most studies looked okay, though a lot of them had pretty small groups. One thing that definitely skews things - if your experiment works, you publish it. If nothing happens, nobody cares. So we're probably seeing better results than what actually happens in real life.

Results and Discussion

Main question: you teach English vocabulary using ears, eyes, and movement all together. Compare that to the usual way - students repeat words or flip through flashcards. Does using multiple senses actually make words stick in memory longer? Not just doing better on tomorrow's quiz. We're talking about whether those words are still there weeks or months later.

Andrä's group (2020) tested this with 8-year-olds learning foreign words. Kids got split into three groups: audio-only (just heard the words with translations), audio-plus-pictures (heard words while looking at pictures), or audio-plus-pictures-plus-gestures (heard words, saw pictures, did hand movements showing what the word meant). Every group got the same amount of time, same words, same number of repetitions. Right after learning, all three groups scored about the same on the vocabulary test - somewhere around 75-80% correct. No real difference yet.

The researchers brought kids back for testing 2 months later. Audio-only group had dropped to 35% correct. Audio-plus-pictures: 52% correct. Audio-plus-pictures-plus-gestures: 71% correct. The multisensory group retained vocabulary nearly as well after 2 months as they did immediately after learning. Another test 6 months post-learning showed similar pattern: multisensory group still remembering most vocabulary, other groups showing substantial forgetting.

This wasn't subtle effect. Adding pictures to audio helped some compared to audio alone. But adding gestures on top of audio and pictures helped way more. The gesture component seemed to be the critical addition creating durable memory. Why? The researchers measured neural activity during learning using fMRI and found gestures activated motor cortex, premotor areas, and created stronger coupling between sensory regions and motor systems. This distributed activation pattern during encoding predicted which words kids remembered months later.

Macedonia and colleagues (2019) found similar results with adults learning vocabulary in unfamiliar language. They compared four learning conditions: reading words silently, reading aloud, reading aloud while watching video of someone performing gesture, reading aloud while watching video AND performing gesture themselves. Testing immediately after showed modest differences - everyone did okay. Testing 2 weeks later showed divergence. Silent reading group forgot most words. Reading aloud helped some. Watching gestures helped more. But performing gestures yourself beat everything else, with retention rates double the silent reading condition.

Critical detail: gestures had to be iconic (look like what they represent). When researchers included arbitrary movements that didn't meaningfully relate to word meanings, the memory benefit disappeared. Random

hand-waving while hearing vocabulary provided no advantage over audio alone. The movements have to engage meaning-related motor representations, not just add motor activity.

Zappa's team (2024) went beyond asking whether multisensory learning helps retention - they wanted to know what kind of neural representations it builds. Taught participants action verbs like push, pull, kick, throw. Ось переписаний фрагмент з унікальною структурою:

Researchers split participants down the middle. First group: conventional teaching-listen, repeat, memorize. Second group: physical execution-participants actually kicked, pushed, threw while simultaneously processing the corresponding vocabulary. Weeks passed. Testing day arrived with brain scanners tracking neural activity during word recall.

Results showed something weird. People from the movement-training group - when they saw the word "kick," the part of their brain that controls leg movement lit up. Not because they were imagining kicking or planning to kick. Just reading the word triggered a motor response. The people who learned the normal way? Their motor systems did nothing during the same vocabulary tasks.

This tells us something important about how multisensory learning actually works. We're not just talking about stronger normal memories. The structure is different. Vocabulary gets wired into physical experience instead of sitting there as isolated words. It's like creating multiple paths to the same information. When you try to remember "kick" months later, you're not just digging through verbal memory. Your brain pulls from: what it feels like when your leg extends, what it looks like when a foot hits something, how your balance shifts. Three different routes to one word. That makes it way harder to forget than words you learned by just repeating them.

Retention's one thing. But you're learning a language to actually use it - speaking and writing. Does multisensory learning help with that, or does it just make you better at recognizing words without being able to produce them?

Macedonia and Repetto (2022) went through studies tackling this question head-on. Across multiple experiments, learners acquiring vocabulary through multisensory methods - especially gesture-enriched learning - retrieved words faster, pronounced them more accurately, used them spontaneously in conversation more often than traditional learning groups. Advantage showed up strongest in early production attempts, first few times trying to actually use new words, suggesting multisensory encoding makes words easier to grab initially.

García-Gámez and Macizo (2023) tested this using gestures as scaffolding for vocabulary production. Spanish speakers learning English vocabulary got divided: some did standard flashcard learning, others performed iconic gestures while studying. Both groups studied until recognizing all words at 90% accuracy. Then production test: researchers showed pictures, participants had to say English words. Gesture learning group spit out words 18% faster on average, 22% more accurately than flashcard group. Gesture training didn't just help recognition - it helped actual production.

Why does this work? One explanation: performing gestures during learning creates motor programs associated with words. Later during production, initiating word production partially reactivates those motor programs, providing retrieval support. Another explanation: gestures during learning draw attention to semantic features, creating richer semantic representations that are easier to access during production attempts. Both probably happen simultaneously.

Lopez-Ozieblo (2024) tested whether gestures help learning abstract grammar, not just concrete nouns and verbs. Taught modal verbs - can, should, must - to young adults. Half got standard explanation, half got gestures added. Gestures showed abstract meaning metaphorically: hand rising progressively for obligation increasing from "can" through "should" to "must." Production tests afterward: gesture group built sentences with modals more accurately, confused different modals less often. Suggests embodied learning works for grammar too when gestures capture abstract relationships meaningfully.

Some new research looked at using VR for language learning. VR gives you control over all the senses at once - you can show objects, play sounds from different directions, track how people move, add touch feedback - stuff that's hard to pull off in a regular classroom.

Elhambakhsh's group (2024) asked English teachers what they thought about using VR. Teachers were split. Yeah, they could see how VR might work - stick students in an immersive environment where they interact with

virtual objects while hearing and speaking English. But most teachers admitted they had no idea how to actually set that up. They don't know how to design VR lessons or how to structure multisensory input in a VR space.

Repetto's group (2016) showed VR-based embodied learning boosted episodic memory in elderly adults. Participants explored virtual environments learning vocabulary for objects and locations they encountered. Multisensory VR condition - visual plus spatial audio plus moving through environment - built stronger memories than desktop condition showing same environments without embodied navigation. Effect hit hardest for learners starting with weaker memory, suggesting multisensory scaffolding helps strugglers most.

The VR research is promising but preliminary. Most studies focus on demonstrating VR can support multisensory learning rather than systematically comparing VR-based versus traditional classroom-based multisensory approaches. We don't yet know whether VR's advantages justify costs and complexity for typical language classrooms. But VR provides research tool for precisely controlling multisensory input combinations and timing, which will help identify optimal implementation parameters.

Does age matter for how well multisensory learning works? Pretty important question for actually using this - if it only helps kids or only helps adults, that changes what teachers should do.

Multisensory stuff works at any age, but probably for different reasons depending on how old you are. Little kids, maybe 5 to 10 years old, respond really well to learning with gestures. Makes sense - they're still figuring out abstract thinking, so they need concrete physical anchors (Macedonia, 2019). Teenagers and young adults benefit too, but not as dramatically as younger kids. Older adults, 60 and up, show strong benefits again, especially for remembering things long-term. Probably because multisensory encoding helps make up for memory systems that start wearing down with age (Repetto et al., 2016).

Esplendori's team (2022) looked at multisensory methods in nursing students - not language learning exactly, but close enough to matter. They found that using multiple senses helped students learn better and stay engaged, even though these were young adults. Point is, multisensory benefits for adults aren't just compensation for getting older. It's how human brains work.

Bottom line: multisensory learning isn't just for kids.

Knowing multisensory learning helps is one thing. Making it work in practice is completely different. Research shows a few key factors that decide whether this approach works or crashes hard.

First, gestures have to mean something. They need to actually represent the word, not just be random arm-waving. Gestures that look like what they're describing, gestures that physically show abstract ideas, gestures that point at things - those work. Random movements do nothing and probably just distract people (Macedonia et al., 2019). Teachers need training to figure out which gestures make sense for the vocabulary they're teaching.

Self-generation versus observation: Learners making gestures themselves creates stronger effects than watching someone else make same gestures. Self-generated movement engages motor systems deeper and builds stronger motor memory traces. But observing gestures still helps versus no gestures, so even if classroom constraints prevent everyone moving simultaneously, teacher modeling gestures gives some benefit (García-Gámez & Macizo, 2023).

Temporal synchrony: Different sensory inputs need coordination. Hearing word simultaneously with seeing picture and making gesture works better than sequential presentation where inputs arrive at different times. Synchronized multisensory input promotes binding into integrated memory trace. Mismatched timing can actually interfere by confusing which inputs belong together (Shams & Seitz, 2008).

Multisensory learning doesn't mean you can skip repetition. Students still need to see and use vocabulary multiple times with the sensory support. The difference is they need fewer repetitions to remember the same amount compared to traditional methods. One study showed multisensory learners needed about 60% as many repetitions as regular learners to remember words equally well six months later (Andrä et al., 2020).

Do you need to explain to students why you're using specific gestures, or do they just pick it up automatically? Depends. Teenagers and adults often benefit from understanding the connection between a movement and what a word means, especially when you're dealing with abstract stuff where the link isn't obvious just by looking.

Little kids though? They don't need the explanation. Their brains just make the connection on their own (Macedonia, 2019).

Now let's get real. Multisensory methods have actual problems that people who love this approach sometimes ignore or brush off. There are some serious gaps worth talking about.

Most research tests concrete vocabulary - nouns for physical stuff, verbs for visible actions. Abstract vocabulary, function words, complex grammar? Way less evidence there. Lopez-Ozieblo's modal verb work suggests embodied approaches could work for grammar, but calling that settled before way more research happens would be premature (Lopez-Ozieblo, 2024).

Nobody knows how much is the right amount. How much multisensory input actually works best? Can you do too much? Does it stop helping at some point? We have no idea. How many gesture-based lessons give you the best results? Can you overdo it in the classroom? Does it level off somewhere, or does more always mean better? Nobody's figured this out. Teachers are just guessing - trying to balance multisensory techniques without boring students or burning them out.

Time's a big unknown. Most studies track students for six months, maybe eight, then stop. What happens at the two-year mark? Four years? Do multisensory learners stay ahead forever, or does everyone end up forgetting at the same rate eventually regardless of how they learned? Nobody knows. We're making guesses.

And here's another thing - labs aren't real life. In a lab, you test vocabulary with neat, controlled prompts and structured responses. No pressure, no distractions. Real language use? Conversations that go off in random directions, giving presentations when you're nervous and can't think straight, writing under deadline when you need exactly the right word and can't find it. Does multisensory learning still help in those messy real situations? Makes sense that it would, and early signs look okay, but nobody's actually tested it properly. We're assuming it transfers without checking.

Discussion

Looking at 15 years of research, multisensory learning in English holds up pretty well. Combining audio, visuals, and movement creates vocabulary memories that last longer than traditional single-channel teaching. Helps with pulling words out when you're speaking. Works across different ages and situations. How big the effect is depends on exactly how you do it, but generally you're looking at moderate to large improvements in retention.

Fits with what we know about embodied cognition and multisensory memory integration. Human brains never evolved processing language as abstract symbols floating separate from sensory and motor experience. Language understanding and production engage distributed neural systems including sensory and motor cortex. Learning engaging these systems builds richer, easier-to-access memory.

What does this mean for actual teaching? Pretty straightforward. Teachers should pair new vocabulary with gestures and pictures every time, not just once in a while as something extra. Doesn't mean you throw out everything else - pronunciation practice, reading, writing, conversation are still important. Just means you build multisensory stuff into regular teaching instead of treating it like a bonus activity.

Here's what that looks like: Teaching concrete nouns like "table" or "dog"? Show a picture while saying the word while doing a gesture that shows the object or what you do with it. Teaching action verbs like "run" or "jump"? Have students actually do the action while saying and hearing the word. Teaching abstract concepts like "freedom" or "justice"? Come up with gestures that represent the meaning somehow, have students do those gestures while they study. Use the multisensory approach multiple times across different lessons. Test students after some time passes, not just right away - that's when multisensory advantages show up strongest.

Tech can help, sure. You can show videos of gestures for students to copy. There's software that puts audio, visuals, and movement together. VR could work for creating immersive multisensory environments, but right now it's too expensive and complicated for most schools. Thing is, you don't need any of that. Teacher shows a gesture, students do it while they study, you throw up some pictures - that's it. No special equipment.

Teacher training programs need to catch up. Most of them don't teach embodied or multisensory methods beyond some vague mention of "kinesthetic learners." Teachers need real training on this stuff: how to pick or create gestures that make sense for vocabulary, how to time multiple sensory inputs so they work together, when to model gestures versus letting students come up with

their own, how to handle abstract concepts. This shouldn't be sold as some trendy new pedagogy. It's just applying what we know about how memory and learning work.

Let's be honest about the problems. Multisensory teaching takes more prep time than traditional methods. You can't just open a textbook and go. You need to develop a whole set of gestures, coordinate materials, plan movement activities. The upfront work pays off because students learn better, but it costs teachers real time and effort. Schools need to support teachers with resources and planning time, not just tell them to do something new without giving them what they need.

Second, managing a classroom gets harder when students are moving around doing gestures instead of sitting still at their desks. Doesn't have to turn into chaos - you can structure gesture activities so they stay controlled - but it takes different skills than traditional teaching. Some teachers will love this. Others will hate it. You can't force everyone to teach the same way, but you also can't ignore what the research shows works.

Third, assessment should align with instruction. Students learning vocabulary through multisensory methods need testing allowing them demonstrating knowledge through multiple modalities, not just traditional paper tests. Might mean production tasks, gesture-based assessments, multimedia demonstrations. Assessment reform complicated enough without adding this but misalignment between instruction and assessment undermines both.

Let's talk about what we don't know. Most studies test vocabulary in labs over a few weeks or months. Real classrooms are nothing like that. You're dealing with thirty different kids, curriculum deadlines, limited resources, and trying to achieve multiple goals at once. Labs give you perfect conditions - materials presented in ideal order, no distractions, students who cooperate. Real classrooms? It's controlled chaos mixed with pressure to cover everything and students with completely different needs and personalities. Will what works in a lab actually survive when you transplant it into that mess? Early signs look good (Gebrekal, 2023), but we're nowhere near sure yet.

Here's another huge problem: almost every study tests concrete nouns and action verbs. Apples. Jumping. Tables. Stuff you can show in a picture or act out. But what about prepositions? What about verb tenses that express hypothetical situations? What about words that show how arguments connect in writing? Lopez-Ozieblo showed some promise using metaphorical gestures for modal verbs, but that's one study on one tiny slice of grammar. Taking results from "kick" and "apple" and assuming they apply to all the complex abstract parts of language? That's a guess, not science.

VR gets way too much hype compared to actual proof it works. Yeah, in theory VR lets you perfectly coordinate audio, visual, touch, and movement in ways you can't do in a regular classroom. Reality? The equipment costs too much. Most schools don't have the tech infrastructure or people who know how to use it. Most VR content is designed for entertainment, not teaching. And nobody's figured out the best way to actually use it for education yet. Maybe VR will be useful in ten years. Right now? Low-tech gestures work and don't require massive budgets.

What we really need: studies that follow students for years, not months, to see if the advantages stick around or disappear. Studies that test abstract grammar systematically instead of just picking easy examples. Research that watches real teachers try this in actual classrooms, sees where it falls apart under real-world pressure, figures out which students it helps and which it doesn't. Studies that calculate the real costs versus the real benefits to see if it's worth the investment. That kind of research would turn this from an interesting idea into something schools can actually use.

Bigger picture: multisensory research challenges assumptions about language and learning that have been around forever. For decades, linguistics and education treated language as a purely mental, abstract system that has nothing to do with physical experience. Teachers learned to treat language as something happening in the mind, separate from the body. So classrooms focused on reading, writing, listening to lectures - activities that minimize physical involvement.

Embodied cognition research pushes back against that split. Mind and body aren't separate things where one does the real learning and the other just carries it around. Thinking, learning, remembering - all of it involves sensory and motor systems as central parts, not extras on the side. For teaching, this means physical engagement isn't a break from learning or a way to motivate students or something you do to accommodate "kinesthetic learners." Physical engagement IS learning when you set it up right.

This doesn't mean we go back to purely hands-on learning or avoid abstract thinking. Language naturally includes symbols and ideas that go beyond what you can see, hear, or touch right now. But those abstract skills don't develop in a vacuum - they grow from physical and sensory experience. Learning works best when teachers help students connect abstract language to real, concrete experiences. Multisensory support is one practical way to do that. It's not a magic solution to every problem in language learning. It can't replace good teaching, clear explanations, practice, or chances to actually use the language. But it's backed by research, most teachers can use it without massive extra effort, and it makes a real difference in how well students remember what they learn.

Conclusion

So does combining sound, visuals, and movement actually help English learners compared to the usual methods? Based on the last 15 years of research - yeah, it does. Effect sizes are moderate to large for remembering vocabulary, smaller but consistent for actually using words. The real advantage shows up when you test students weeks or months later - multisensory learning creates memories that last way longer than just hearing words or seeing flashcards. Works across ages too, from little kids to older adults.

What makes it work: gestures have to actually mean something (random movements do nothing), students need to do the gestures themselves (watching doesn't cut it), everything needs to happen at the same time (don't separate the audio from the visual from the movement), and you still need repetition (multisensory cuts down how much practice you need but doesn't eliminate it).

For teachers: build gestures and visuals into how you teach new vocabulary every time, not when you happen to have extra time. Have students do the gestures themselves, don't just demonstrate while they watch. Present words, pictures, and movements together so students can connect them. Test what students remember after time passes, not just the next day - that's when you'll see what actually stuck. Even abstract concepts can work this way if you use gestures that represent ideas metaphorically.

But let's be clear about what we don't know. Most research focuses on concrete words - apples, jumping, tables. How well does this work for abstract language? We're guessing. Nobody knows the optimal dose - how much multisensory input, how often. Long-term data barely exists since most studies stop at a few months. New tech like VR sounds exciting but isn't practical for most classrooms yet. And teacher training programs mostly ignore this stuff, so teachers are left to figure it out on their own.

Why this matters beyond just classroom tricks: it supports the idea that language learning is tied to the body and senses, not just memorizing abstract symbols in your head. That challenges traditional assumptions about how languages should be taught and suggests we might need to rethink teaching methods more broadly.

Multisensory learning isn't a magic fix. It won't replace good teaching, clear explanations, practice, or real chances to communicate. But it's one practical tool backed by solid research that most teachers can use without needing special equipment or massive prep time. And when used right, it makes a real difference in how well students learn and remember.

References

1. Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & von Kriegstein, K. (2020). Learning foreign language vocabulary with gestures and pictures enhances vocabulary memory for several months post-learning in eight-year-old school children. Educational Psychology Review, 32(3), 815-850. https://doi.org/10.1007/s10648-020-09527-z
2. Elhambakhsh, S. E., Neysani, M., & Nikbakht, A. (2024). Exploring L2 educators' training and professional development needs in VR English language learning. Heliyon, 10(17), e36700. https://doi.org/10.1016/j.heliyon.2024.e36700
3. Esplendori, G. F., Kobayashi, R. M., & Püschel, V. A. A. (2022). Multisensory integration approach, cognitive domains, meaningful learning: reflections for undergraduate nursing education. Revista da Escola de Enfermagem da USP, 56, e20210381. https://doi.org/10.1590/1980-220X-REEUSP-2021-0381
4. Fallace, T. (2023). The long origins of the visual, auditory, and kinesthetic learning style typology, 1921-2001. History of Psychology, 26(4), 334-354. https://doi.org/10.1037/hop0000240
5. García-Gámez, A. B., & Macizo, P. (2023). Gestures as scaffolding to learn vocabulary in a foreign language. Brain Sciences, 13(12), 1712. https://doi.org/10.3390/brainsci13121712

6. Gebrekal, G. A. (2023). Accommodating students' learning styles differences in English language classroom. Heliyon, 9(6), e17497. https://doi.org/10.1016/j.heliyon.2023.e17497
7. Lopez-Ozieblo, R. (2024). Effects of pedagogical gestures on learning abstract grammatical concepts in young adults. Frontiers in Communication, 9, 1372033. https://doi.org/10.3389/fcomm.2024.1372033
8. Macedonia, M. (2019). Embodied learning: Why at school the mind needs the body. Frontiers in Psychology, 10, 2098. https://doi.org/10.3389/fpsyg.2019.02098
9. Macedonia, M., & Repetto, C. (2022). Embodied learning and teaching approaches in language education: A mixed studies review. Educational Research Review, 37, 100480. https://doi.org/10.1016/j.edurev.2022.100480
10. Macedonia, M., Repetto, C., Ischebeck, A., & Mueller, K. (2019). Depth of encoding through observed gestures in foreign language word learning. Frontiers in Psychology, 10, 33. https://doi.org/10.3389/fpsyg.2019.00033
11. Pelkey, J. (2023). Embodiment and language. WIREs Cognitive Science, 14(5), e1649. https://doi.org/10.1002/wcs.1649
12. Reggin, L. D., Gómez Franco, L. E., Horchak, O. V., Labrecque, D., Lana, N., Rio, L., & Vigliocco, G. (2023). Consensus paper: Situated and embodied language acquisition. Journal of Cognition, 6(1), 63. https://doi.org/10.5334/joc.308
13. Repetto, C., Serino, S., Macedonia, M., & Riva, G. (2016). Virtual reality as an embodied tool to enhance episodic memory in elderly. Frontiers in Psychology, 7, 1839. https://doi.org/10.3389/fpsyg.2016.01839
14. Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. Trends in Cognitive Sciences, 12(11), 411-417. https://doi.org/10.1016/j.tics.2008.07.006
15. Zappa, A., Bolger, D., Pergandi, J. M., Fargier, R., Mestre, D., & Frenck-Mestre, C. (2024). The neural correlates of embodied L2 learning: Does embodied L2 verb learning affect representation and retention? Neurobiology of Language, 5(2), 360-384. https://doi.org/10.1162/nol_a_00132