

Organizing Learner Diversity in Mixed-Ability Foreign Language Classrooms: A Critical Review and Framework for Tiered Task Design

Bin Wang*

Shanghai Aurora College, Shanghai 201908, China

*Corresponding author: Bin Wang, b.wang@aurora-college.cn

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Learner diversity in foreign language classrooms is often handled as a placement problem. Students are grouped by entrance or proficiency scores, and differentiated instruction is expected to follow from the resulting class structure. This critical integrative review questions that assumption and develops a framework for tiered task design in mixed-ability foreign language classrooms. Drawing on research on ability grouping, tracking, differentiated instruction, second language motivation, task-based language teaching, dynamic assessment, and recent AI-mediated EFL support, the article argues that the central issue is not whether learners differ, but how those differences are identified, interpreted, adjusted for, and made consequential in classroom life. Evidence from general education offers limited support for long-term fixed between-class grouping and points to risks of labeling, reduced teacher expectations, weakened academic self-concept, and unequal access to demanding tasks. Evidence from differentiated instruction and EFL task studies is more favorable to within-class adjustment, although it remains fragmented and does not yet establish tiered task design as a superior intervention. On this basis, the review proposes a four-dimensional account of learner-diversity organization and specifies four mechanisms for tiered task design: shared communicative goals, graded scaffolding, scaffolded choice, and developmental assessment. The framework is presented as a researchable pedagogical model rather than a settled solution, with implications for language programs seeking to support participation and equity without turning proficiency differences into fixed learner labels.

Keywords: Ability grouping; Differentiated instruction; EFL; Learner diversity; Mixed-ability classrooms; Task-based teaching; Tiered tasks

Online publication: August 4, 2026

1. Introduction

Mixed-ability teaching is not an exceptional circumstance in foreign language education. It is closer to the ordinary condition under which much classroom language learning takes place. Learners enter the same

course with uneven histories of exposure, vocabulary knowledge, grammatical control, literacy practices, willingness to communicate, and access to learning resources. These differences are not merely individual variables that teachers must accommodate after instruction has been planned. They are also produced and stabilized by institutional arrangements, including placement tests, course labels, classroom grouping, materials, assessment regimes, and the expectations attached to each of these practices. A language program therefore does more than observe learner diversity; it organizes it.

A common organizational response is ability grouping. In many programs, learners are assigned to different classes on the basis of entrance examinations, placement tests, or teacher judgments of proficiency. The rationale is straightforward: if classes are more homogeneous, teachers can select a common pace, adjust materials more efficiently, and avoid aiming instruction either too high or too low. The appeal of this arrangement is practical, and it should not be dismissed as simply misguided. In large programs with limited staffing, broad placement categories may appear to be the only administratively manageable response to heterogeneous proficiency.

The difficulty is that placement arrangements do not only manage instruction. They also create categories through which learners and teachers interpret ability. Once a learner has been placed in a low-level class, the placement may come to function as a durable identity marker rather than a temporary pedagogical decision. In foreign language education, this matters because proficiency is not a single trait. A learner who struggles with written accuracy may participate well in oral rehearsal; a learner with a weak test score may have strong domain knowledge, digital literacy, or interactional initiative. If a program treats a prior score as a general description of language-learning capacity, it risks narrowing both instruction and learner possibility.

This review takes up that problem by asking how mixed-ability foreign language classrooms might organize learner diversity without converting it into fixed learner labels. The focus is not on denying proficiency differences, nor on romanticizing heterogeneity. The question is more specific: when difference is present, should the primary pedagogical unit of differentiation be the learner, the class, or the task? The review develops the argument that fixed learner-based grouping and dynamic task-based differentiation represent different ways of making learner difference consequential. The latter does not remove the need for diagnosis, support, or challenge; rather, it relocates differentiation from class membership to task complexity, scaffolding, choice, interaction, and assessment evidence.

The article makes three contributions. It brings research on ability grouping, differentiated instruction, L2 motivation, task-based language teaching, and dynamic assessment into conversation around a problem that is often treated as a local management issue. It develops a four-dimensional account of learner-diversity organization: identifying diversity, interpreting diversity, adjusting instruction, and shaping consequences. It then proposes a task-tiering framework, especially relevant for mixed-ability and task-rich EFL contexts, that connects shared communicative goals with graded scaffolding, scaffolded choice, and developmental assessment. The framework is deliberately presented as a researchable model rather than as proof that tiered tasks are universally superior to grouped classes. That caution is necessary because the strongest evidence on ability grouping comes largely from general education, whereas direct evidence on tiered tasks in EFL remains modest.

2. Review scope and conceptual clarifications

This article is a critical integrative review rather than a meta-analysis or a PRISMA-style systematic review.

Its purpose is to synthesize adjacent literatures that usually remain separate: general education studies of ability grouping and tracking; research on differentiated instruction and instructional adaptation; work on academic self-concept, teacher expectations, and motivation; L2 task-based teaching and dynamic assessment; and a smaller body of EFL research on tiered tasks and AI-mediated support. The review therefore prioritizes conceptual fit and evidentiary relevance over exhaustive database coverage.

The literature was read with two guiding questions in mind. One concerns the consequences of grouping learners by prior achievement: What does the evidence suggest about the academic, motivational, and equity effects of fixed grouping? The other concerns within-class alternatives: Under what conditions might task-based differentiation offer a more developmentally responsive way to support heterogeneous learners? Because the language education literature has changed rapidly since the public release of large language models, recent work on GenAI-mediated EFL learning published between 2024 and 2025 is also considered. These studies are not treated as direct evidence that tiered tasks work. They are used more narrowly to examine how newer forms of diagnostic feedback, adaptive support, and teacher-student-AI interaction complicate the problem of organizing learner diversity ^[1,2]. See **Table 1**.

Several terms require clarification. Ability grouping refers broadly to grouping students by prior achievement or perceived ability. Between-class ability grouping places learners into different classes, usually for a term, year, or program sequence. Within-class grouping organizes learners temporarily inside the same class. Setting is often subject-specific, whereas tracking and streaming may involve longer-term curricular pathways and more consequential institutional sorting. These distinctions are not merely terminological. Evidence on tracking cannot be transferred without qualification to language-course placement, and findings on short-term within-class grouping do not automatically validate any form of task tiering.

In many EFL programs, level-based placement is closest to between-class grouping or subject-specific setting. By contrast, tiered task design retains a common class and common instructional goal while varying the route into the task. Learners may receive different scaffolds, planning time, models, interactional roles, or output requirements, but they remain oriented to a shared communicative purpose. This distinction is central to the argument. The issue is not whether differentiation occurs; it is whether differentiation is attached primarily to learners as categories or to tasks as adjustable pedagogical environments.

Table 1. Related practices and their relevance to the review

Practice	Primary object	Typical duration	Relevance to this review
Tracking / streaming	Educational pathway or curriculum route	Long-term	Shows how selection can affect opportunity and identity
Between-class ability grouping	Class membership	Term or academic year	Closest to many level-based language programs
Setting	Subject-specific grouping	Variable	Shows that grouping may be local to one subject
Within-class grouping	Temporary group or task role	Short-term	Provides a route to flexible differentiation
Tiered tasks	Task pathway, support, and output demand	Lesson, unit, or project	Central alternative developed in this review

3. Fixed learner-based grouping: evidence and risks

Ability grouping has a long empirical history, and the findings are more qualified than either advocates or

critics sometimes suggest. Slavin's best-evidence syntheses found little support for self-contained between-class grouping in elementary and secondary education, while reporting more favorable results for more flexible arrangements such as within-class or cross-grade subject grouping ^[3,4]. Steenbergen-Hu *et al.* ^[5], in a second-order meta-analysis of a century of research on ability grouping and acceleration, likewise distinguish among grouping forms. Their synthesis suggests that between-class grouping produces only small average achievement effects, whereas within-class grouping, cross-grade subject grouping, and special grouping for high-achieving students show stronger benefits. The finding does not imply that all grouping is harmful. It suggests that the pedagogical consequences of grouping depend on mobility, curriculum quality, and the learning opportunities that grouping makes available.

Recent evidence on tracking points to a related concern. Terrin and Triventi ^[6] reported that school tracking has no reliable positive effect on average achievement but is associated with greater inequality. OECD ^[7] similarly connected earlier or stronger forms of selection with wider achievement and socioeconomic gaps across systems. These studies largely concern general education and often K-12 contexts, so they cannot be read as direct evidence against EFL placement classes. Yet they are relevant because they expose a mechanism that language programs should not ignore: selection arrangements can distribute not only learners but also expectations, curricula, peer networks, and access to demanding work.

Psychological mechanisms help explain why fixed grouping may have consequences beyond scheduling. When learners are placed in lower groups, they may come to treat the placement as evidence of durable language-learning capacity rather than as a provisional instructional arrangement. Francis *et al.* ^[8], for example, showed that attainment grouping can have cumulative effects on pupils' self-confidence even after prior achievement is taken into account. Merton's ^[9] account of the self-fulfilling prophecy and the classic work of Rosenthal and Jacobson ^[10] on teacher expectations have been debated and refined, but the broader literature continues to show that expectations are linked to students' opportunities and outcomes. Wang *et al.* ^[11] synthesized evidence that expectation effects are not evenly distributed; they matter more for learners already positioned as low-achieving or stigmatized. In a language classroom, lower expectations may appear in subtle forms: fewer open-ended questions, less wait time, less extended output, more correction without elaborated feedback, or tasks that reward avoidance of error more than communicative risk-taking.

Research on academic self-concept complicates the picture further. The big-fish-little-pond effect indicates that students may judge their ability partly in relation to the average achievement of the group in which they are placed ^[12,13]. This finding is sometimes used to defend lower-level classes on the grounds that weaker learners may experience less threatening comparison. The inference is too quick. Self-concept is not the same as access to rich input, challenging tasks, or capable peers. A lower-ability group may reduce one form of comparison pressure while also reducing opportunities for upward modeling and interaction. Roy *et al.* ^[14] suggested that differentiated instruction can moderate such comparison effects, which is precisely why the issue should be framed not as grouping versus no grouping but as the design of learning conditions inside heterogeneous settings.

4. A four-dimensional account of learner-diversity organization

The previous section suggests that fixed grouping should not be evaluated only by average test-score effects. It should also be examined as a way of organizing learner diversity. Learner-diversity organization is used to refer to the processes through which language programs identify differences among learners, interpret

those differences pedagogically, adjust instruction in response, and shape the consequences that follow. This framing helps explain why two approaches that both claim to differentiate instruction may have quite different effects.

The first dimension is the identification of diversity. Placement systems usually begin with prior scores or one-off tests. Such evidence may be useful, especially for broad program placement, but it captures only a narrow slice of language learning. It says little about genre-specific performance, interactional confidence, vocabulary access under time pressure, multimodal composing, or the kinds of support that might allow a learner to participate in a particular task. Task-based differentiation requires a more local diagnostic question: What support does this learner need to work productively in this task? Dynamic assessment is useful here because it treats assessment as a means of revealing modifiability, not simply independent performance^[15].

Recent AI research in EFL gives this issue a new form but not a simple solution. Scoping and systematic reviews suggest that AI tools can expand feedback, content generation, and individualized practice, while also raising problems of teacher readiness, digital inequality, cognitive load, privacy, and context-sensitive adaptation^[1,2]. For a tiered-task classroom, the implication is not that diagnostic work can be outsourced to technology. It is that diagnostic evidence must remain pedagogically interpreted. Teachers still need to decide whether a learner needs lexical support, rehearsal time, interactional modeling, a more demanding communicative problem, or simply a clearer criterion for revision.

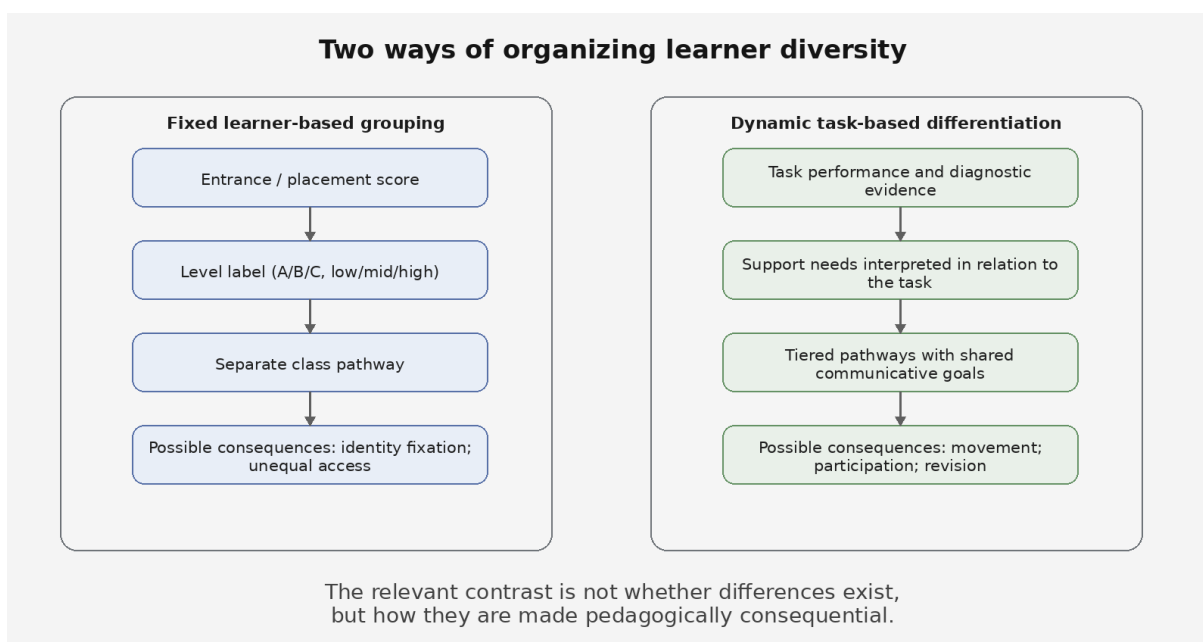
The second dimension is interpretation. Fixed grouping often encourages differences to be interpreted as stable levels: high, middle, low. Task-based differentiation interprets differences as relations among learner, task, support, and context. The same learner may require strong linguistic scaffolding for an argumentative writing task but much less support for a role-play based on domain knowledge. This view is consistent with Vygotsky's^[16] account of development as mediated activity. It also fits language classroom realities better than a single proficiency label, since performance varies across genres, modalities, audiences, and interactional demands.

The third dimension is adjustment. Fixed grouping adjusts instruction through separation: different classes, different pace, different materials, sometimes different assessments. Task tiering adjusts instruction inside a common task ecology. Complexity, support, interactional roles, and assessment evidence are varied while the communicative purpose remains shared. This position is close to Tomlinson's^[17] account of differentiated instruction, but language teaching requires a more specific rationale. Task-based language teaching provides one by showing how tasks can be designed and sequenced through cognitive demands, planning time, information distribution, reasoning requirements, modality, and focus on form^[18–20].

The final dimension concerns consequences. Fixed grouping may create durable learner identities and unequal access to demanding work. Task tiering aims for movement: learners enter through different pathways but are not meant to remain in them. The intended consequence is not a more precise sorting system, but visible development across attempts, revisions, and increasingly demanding participation. **Table 2** summarizes this contrast. **Figure 1** shows the two ways of organizing learner diversity in mixed-ability foreign language classrooms.

Table 2. Four dimensions of learner-diversity organization

Dimension	Fixed learner-based grouping	Dynamic task-based differentiation
Identifying diversity	Prior scores or one-off placement evidence	Task performance, diagnostic evidence, and support needs
Interpreting diversity	A relatively stable level or category	A revisable relation among learner, task, support, and context
Adjusting instruction	Separate classes, pacing, materials, or tests	Task complexity, scaffolds, interactional roles, and assessment evidence
Shaping consequences	Possible identity fixation and unequal access to demanding work	Potential movement across pathways while retaining shared goals

**Figure 1.** Two ways of organizing learner diversity in mixed-ability foreign language classrooms

5. Mechanisms of tiered task design

A task-tiered approach depends on more than offering easier and harder versions of the same activity. Without a principled design, tiering can become a more polite name for low expectations. Four mechanisms are therefore necessary: shared communicative goals, graded scaffolding, scaffolded choice, and developmental assessment.

Shared communicative goals keep tiering inside a common curriculum. Learners may receive different scaffolds, but they should remain oriented to the same task purpose. If basic-tier learners complete decontextualized worksheets while challenge-tier learners engage in communication, the design reproduces the very division it is meant to avoid. Task-based teaching strengthens this point. Bryfonski and McKay's ^[21] meta-analysis suggests favorable effects for task-based language teaching, but its relevance here is not that tasks are automatically effective. It is that tasks provide a principled way to organize language learning around meaning, interaction, and outcome. Tiering should therefore modify access to the task rather than replace the task with unrelated drills.

Graded scaffolding follows from the same logic. Basic tiers should not be low-expectation spaces. They should offer stronger access supports: vocabulary banks, sentence frames, models, rehearsal sequences, peer prompting, teacher feedback, or AI-assisted practice when appropriate. Standard tiers can represent the expected curricular outcome. Challenge tiers can increase audience awareness, reasoning demands, transfer requirements, multimodal design, or opportunities for linguistic refinement. The difference among tiers should be in support and complexity, not in whether learners are allowed to participate in meaningful communication.

Choice is more difficult. Self-determination theory argues that autonomy, competence, and relatedness support higher-quality motivation ^[22,23], and meta-analytic evidence links autonomous forms of motivation to persistence and positive outcomes ^[24]. Choice can also increase interest and effort under appropriate conditions ^[25]. Yet unrestricted choice may be counterproductive in tiered tasks. Learners with low self-efficacy may repeatedly choose below their potential; others may choose a challenge tier without sufficient support and then withdraw. The problem is not solved by exhortations to develop a growth mindset, especially given the modest average effects reported by Sisk *et al.* ^[26]. A more defensible design is scaffolded choice: teacher diagnostic feedback, a standard pathway as the default curricular expectation, a basic pathway as entry support, and a challenge pathway as a guided extension.

Developmental assessment completes the mechanism. If assessment recognizes only final products, tiering may reward already advantaged learners. If it recognizes only effort, it may blur quality standards. Developmental assessment should instead make movement visible: draft revision, oral rehearsal, feedback uptake, portfolio evidence, peer support, and justified task upgrading. In Chinese EFL contexts, the Production-Oriented Approach offers a useful connection. Wen's ^[27] model of motivating, enabling, and assessing output, together with teacher-student collaborative assessment ^[28], treats assessment as part of learning rather than as a terminal sorting device. This is compatible with tiered tasks, provided that assessment criteria remain common enough to preserve curricular equity while flexible enough to recognize different routes into performance. See **Table 3**. **Figure 2** illustrates the mechanism of tiered task design in mixed-ability foreign language classrooms.

Table 3. Mechanisms of tiered task design

Mechanism	Pedagogical function	Theoretical support	Risk if absent
Shared communicative goals	Keep tiering inside a common curriculum	TBLT; equity-oriented curriculum design	Tiers become separate low- and high-level curricula
Graded scaffolding	Provide access without lowering expectations	ZPD; dynamic assessment; differentiated instruction	Basic tiers become low-expectation worksheets
Scaffolded choice	Connect autonomy with competence support	Self-determination theory; choice research	Learners repeatedly choose too low or too high
Developmental assessment	Make revision and movement visible	POA; teacher-student collaborative assessment	Tiering becomes another form of sorting

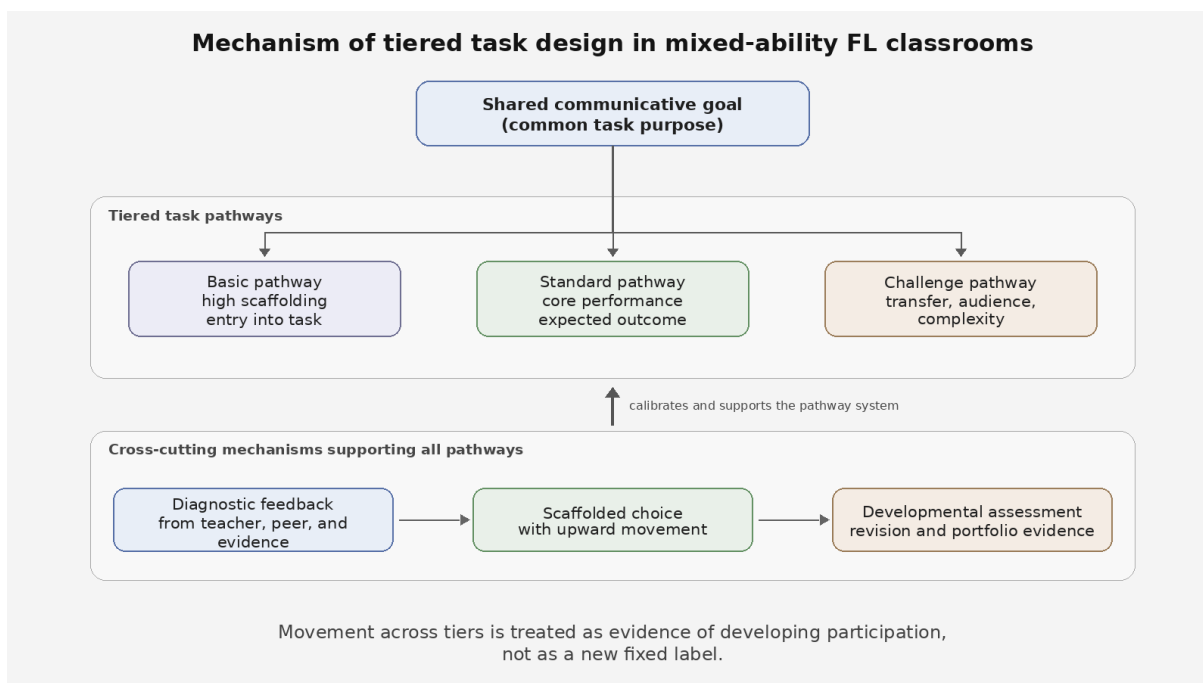


Figure 2. Mechanism of tiered task design in mixed-ability foreign language classrooms

6. Evidence for tiered tasks and its limits

The evidence base for task tiering in EFL is promising but not yet robust. Differentiated instruction research provides indirect support. Smale-Jacobse *et al.* ^[29] reviewed studies in secondary education and report generally positive but methodologically uneven evidence. Deunk *et al.* ^[30] reached a similar conclusion for differentiation practices in primary education: effects appear possible, but they depend on diagnostic quality and instructional adaptation. Parsons *et al.* ^[31] also showed that teachers' instructional adaptations are a recurrent feature of responsive teaching. These findings justify further work on within-class adjustment, but they do not by themselves demonstrate that tiered language tasks outperform level-based placement.

The more directly relevant EFL evidence is smaller. Pourdana and Shahpouri Rad ^[32] implemented tiered listening tasks in a mixed-ability EFL context and reported benefits for listening comprehension, with learners moving across levels according to performance over the intervention. Rafi and Pourdana ^[33] examined e-diagnostic assessment of collaborative and individual oral tiered tasks among non-English-major university students; their findings support the value of combining diagnostic evidence, tiered oral tasks, and collaboration. Sapan and Mede ^[34] reported positive effects of differentiated instruction on achievement, motivation, and autonomy in a school-based EFL setting. These studies are valuable because they move beyond general education and address language learning. They also show the limits of the current evidence: sample sizes are modest, contexts differ from many tertiary EFL programs, and the interventions often involve teacher-assigned or diagnostically adjusted tiers rather than fully learner-selected pathways.

Recent AI-mediated EFL studies add a new but still bounded layer of evidence. Huang and Mizumoto ^[35] reported that generative AI use in an instructor-led EFL writing classroom was associated with changes in components of the L2 motivational self system. Chen *et al.* ^[36] found that role-play with a GenAI agent improved learners' intrinsic motivation and self-efficacy, although speaking-performance gains did not

significantly exceed peer role-play. Guan *et al.* ^[37] argued for a tripartite view of EFL instruction in which teachers, students, and generative AI tools interact rather than technology replacing human mediation. Research on chatbot-supported speaking also suggests possible benefits for anxiety, enjoyment, interactional practice, and performance, while remaining sensitive to task structure and feedback quality ^[36,38].

These AI studies should not be overinterpreted. They do not prove that tiered tasks work, nor do they establish that AI provides reliable differentiation. Their value lies elsewhere: they show that additional interlocutors and feedback systems can change the conditions under which learners rehearse, revise, and participate. For tiered task design, AI-mediated support is best treated as one possible scaffold among others. It remains subordinate to the pedagogical problem of maintaining shared goals across differentiated pathways. A chatbot role-play may support a basic pathway by increasing rehearsal time, or it may enrich a challenge pathway by simulating a more demanding audience. In both cases, its value depends on teacher mediation and on whether learners can connect the AI-mediated practice to the common communicative task. Lamb's ^[39] review of the motivational dimension of language teaching is relevant here: motivation is shaped by classroom tasks, materials, teacher support, and the meanings learners attach to participation, not by technology or grouping arrangements alone.

Cooperative learning offers a further mechanism but also a warning. Roseth *et al.* ^[40] showed that cooperative goal structures can support achievement and peer relationships. In mixed-ability language classrooms, stronger learners may provide models and interactional support, while weaker learners' questions may prompt explanation and reformulation. Yet collaboration can also hide unequal participation. If roles are vague and assessment is purely group-based, stronger learners may complete the task while others remain observers. Tiered task design therefore has to combine heterogeneous collaboration with individual accountability, process evidence, and teacher follow-up.

7. Implications for task-rich EFL programs

The framework is not tied to a single sector of EFL provision. It is most relevant to task-rich programs in which learners move between language practice and communicative performance, including project-based, task-based, content-mediated, and AI-supported EFL courses. The theoretical issue is not the institutional label of the course but the pedagogical ecology it creates: learners must interpret a situation, select language resources, coordinate with others, and revise an outcome. A score-based placement decision captures only part of that ecology.

In these programs, initial placement information may still be useful. The problem begins when a placement score is allowed to settle what learners are presumed capable of doing across tasks. A more cautious use of diagnosis is to let initial evidence inform task design: which pathway should offer more lexical or discourse support, where rehearsal is needed before public performance, and how challenge tasks can add transfer, audience awareness, or multimodal demands.

Recent AI-mediated EFL research adds another layer to this argument. Generative tools and chatbots may help teachers produce examples, feedback, and differentiated practice opportunities, but they can also increase dependency, uneven access, and weak task calibration if treated as substitutes for pedagogical judgment ^[1,2,36–38]. AI-supported differentiation should therefore be evaluated by the quality of the task pathway it helps sustain, not by the mere presence of adaptive feedback.

8. Propositions for future research

The framework developed here is not meant to close the question empirically. It is meant to make the question researchable. Four propositions follow from the review.

Proposition 1: In mixed-ability foreign language classrooms, dynamic task-based differentiation is more likely than fixed learner-based grouping to preserve learners' academic self-concept and classroom belonging, provided that learners remain oriented to shared communicative goals. This proposition links grouping research with academic self-concept and should be tested through designs that measure identity-related outcomes as well as achievement.

Proposition 2: The effect of tiered tasks on engagement is moderated by the quality of scaffolded choice. Completely free choice may lead some learners to choose below their developmental potential, while teacher-imposed levels may reproduce labeling. The empirical issue is not whether learners choose, but how diagnostic feedback, default expectations, and opportunities for upgrading shape those choices over time.

Proposition 3: The language-learning value of tiered tasks depends on alignment among task complexity, language scaffolding, and assessment feedback. Increasing or reducing task difficulty alone is unlikely to be sufficient. Research should examine whether the scaffold provided actually corresponds to the linguistic and interactional demands of the task.

Proposition 4: Heterogeneous collaboration becomes productive only when individual accountability and developmental assessment are built into the design. Without visible evidence of participation and revision, mixed-ability grouping may reproduce unequal participation even inside a seemingly inclusive task structure.

These propositions suggest a research agenda that is methodologically broader than simple comparisons of level-based and mixed-ability classes. Quasi-experimental studies could compare fixed placement, conventional mixed-ability teaching, and task-tiered teaching on oral performance, writing quality, self-efficacy, and participation. Longitudinal studies could track learners' movement across pathways and examine whether basic-pathway participation becomes a bridge to more demanding performance or a new label. Classroom-process studies could examine how teachers interpret diagnostic evidence, calibrate tiers, and manage peer and AI-mediated scaffolds. The recent AI literature also raises a specific question: can teacher-facing analytics and AI-supported practice help teachers calibrate tiers without intensifying surveillance, dependence, or unequal access? That question is especially pressing in resource-constrained EFL settings.

9. Conclusion

Learner diversity cannot be solved by sorting. It can only be organized, and different forms of organization carry different pedagogical and identity consequences. Fixed between-class ability grouping may help programs manage instruction, but the evidence reviewed here gives little reason to treat it as a sufficient response to heterogeneous language classrooms. The risks are not only academic. They include the possibility that prior scores become learner labels, that lower placements receive less demanding work, and that expectations are distributed unevenly.

Tiered task design offers a more cautious alternative. It does not deny proficiency differences, nor does it promise that mixed-ability teaching will automatically produce equity. Its value lies in a shift of pedagogical attention: from assigning learners to levels toward designing task environments in which different levels of support, complexity, and assessment evidence remain connected to common communicative goals. This is a modest claim, but it is also a consequential one. If differentiation is located in tasks rather than in learner

identity, movement becomes possible. Learners can enter with support, revise with evidence, and attempt more demanding participation without being reclassified as a different kind of learner.

The framework advanced in this review should therefore be read as a research agenda rather than an implementation recipe. It asks language educators to examine how they identify learner differences, how they explain those differences, how they adjust instruction, and what consequences follow. For mixed-ability foreign language classrooms, especially in task-rich EFL settings, the most defensible direction is not to replace one sorting mechanism with another, but to build instructional designs in which diversity is pedagogically interpreted, supported, and made revisable.

Disclosure statement

The author declares no conflict of interest.

Declaration of generative AI

During the preparation of this manuscript, the author used generative AI tools to support literature organization and academic language revision. After using these tools, the author reviewed, verified, and edited the manuscript content, including the interpretation of the literature, the accuracy of references, and the coherence of the argument. The author takes full responsibility for the content of the manuscript.

References

- [1] Kundu A, Bej T, 2025, Transforming EFL Teaching with AI: A Systematic Review of Empirical Studies. *International Journal of Artificial Intelligence in Education*, 35: 2281–2314. <https://doi.org/10.1007/s40593-025-00470-0>
- [2] Law L, 2024, Application of Generative Artificial Intelligence (GenAI) in Language Teaching and Learning: A Scoping Literature Review. *Computers and Education Open*, 6: Article 100174. <https://doi.org/10.1016/j.caeo.2024.100174>
- [3] Slavin RE, 1987, Ability Grouping and Student Achievement in Elementary Schools: A Best-Evidence Synthesis. *Review of Educational Research*, 57(3): 293–336. <https://doi.org/10.3102/00346543057003293>
- [4] Slavin RE, 1990, Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis. *Review of Educational Research*, 60(3): 471–499. <https://doi.org/10.3102/00346543060003471>
- [5] Steenbergen-Hu S, Makel MC, Olszewski-Kubilius P, 2016, What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K–12 Students' Academic Achievement: Findings of Two Second-Order Meta-Analyses. *Review of Educational Research*, 86(4): 849–899. <https://doi.org/10.3102/0034654316675417>
- [6] Terrin É, Triventi M, 2023, The Effect of School Tracking on Student Achievement and Inequality: A Meta-Analysis. *Review of Educational Research*, 93(2): 236–274. <https://doi.org/10.3102/00346543221100850>
- [7] OECD, 2023, PISA 2022 Results (Volume II): Learning During—and From—Disruption, OECD Publishing. <https://doi.org/10.1787/a97db61c-en>
- [8] Francis B, Craig N, Hodgen J, et al., 2020, The Impact of Tracking by Attainment on Pupil Self-Confidence Over Time: Demonstrating the Accumulative Impact of Self-Fulfilling Prophecy. *British Journal of Sociology of*

- Education, 41(5): 626–642. <https://doi.org/10.1080/01425692.2020.1763162>
- [9] Merton RK, 1948, The Self-Fulfilling Prophecy. *The Antioch Review*, 8(2): 193–210. <https://doi.org/10.2307/4609267>
- [10] Rosenthal R, Jacobson L, 1968, *Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development*, Holt, Rinehart & Winston.
- [11] Wang S, Rubie-Davies CM, Meissel K, 2018, A Systematic Review of the Teacher Expectation Literature Over the Past 30 Years. *Educational Research and Evaluation*, 24(3–5): 124–179. <https://doi.org/10.1080/13803611.2018.1548798>
- [12] Marsh HW, Hau KT, 2003, Big-Fish-Little-Pond Effect on Academic Self-Concept: A Cross-Cultural Test of the Negative Effects of Academically Selective Schools. *American Psychologist*, 58(5): 364–376. <https://doi.org/10.1037/0003-066X.58.5.364>
- [13] Fang J, Huang X, Zhang M, et al., 2018, The Big-Fish-Little-Pond Effect on Academic Self-Concept: A Meta-Analysis. *Frontiers in Psychology*, 9: Article 1569. <https://doi.org/10.3389/fpsyg.2018.01569>
- [14] Roy A, Guay F, Valois P, 2015, The Big-Fish-Little-Pond Effect on Academic Self-Concept: The Moderating Role of Differentiated Instruction and Individual Achievement. *Learning and Individual Differences*, 42: 110–116. <https://doi.org/10.1016/j.lindif.2015.07.009>
- [15] Poehner ME, 2008, *Dynamic Assessment: A Vygotskian Approach to Understanding and Promoting L2 Development*, Springer.
- [16] Vygotsky LS, 1978, *Mind in Society: The Development of Higher Psychological Processes*, Harvard University Press, Cambridge, MA.
- [17] Tomlinson CA, 2014, *The Differentiated Classroom: Responding to the Needs of All Learners*, 2nd ed., ASCD, Alexandria, VA.
- [18] Ellis R, Skehan P, Li S, et al., 2019, *Task-Based Language Teaching: Theory and Practice*, Cambridge University Press, Cambridge.
- [19] Robinson P, 2001, Task Complexity, Task Difficulty, and Task Production: Exploring Interactions in a Componential Framework. *Applied Linguistics*, 22(1): 27–57. <https://doi.org/10.1093/applin/22.1.27>
- [20] Skehan P, 1998, *A Cognitive Approach to Language Learning*, Oxford University Press, Oxford.
- [21] Bryfonski L, McKay TH, 2019, TBLT Implementation and Evaluation: A Meta-Analysis. *Language Teaching Research*, 23(5): 603–632. <https://doi.org/10.1177/1362168817744389>
- [22] Deci EL, Ryan RM, 2000, The “What” and “Why” of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11(4): 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- [23] Ryan RM, Deci EL, 2017, *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*, Guilford Press, New York.
- [24] Howard JL, Bureau J, Guay F, et al., 2021, Student Motivation and Associated Outcomes: A Meta-Analysis from Self-Determination Theory. *Perspectives on Psychological Science*, 16(6): 1300–1323. <https://doi.org/10.1177/1745691620966789>
- [25] Patail EA, Cooper H, Robinson JC, 2008, The Effects of Choice on Intrinsic Motivation and Related Outcomes: A Meta-Analysis of Research Findings. *Psychological Bulletin*, 134(2): 270–300. <https://doi.org/10.1037/0033-2909.134.2.270>
- [26] Sisk VF, Burgoyne AP, Sun J, et al., 2018, To What Extent and Under Which Circumstances Are Growth Mind-Sets Important to Academic Achievement? Two Meta-Analyses. *Psychological Science*, 29(4): 549–571. <https://doi.org/10.1177/0956797618761111>

doi.org/10.1177/0956797617739704

- [27] Wen Q, 2015, Developing a Theoretical System of the Production-Oriented Approach. *Foreign Language Teaching and Research*, 47(4): 547–558.
- [28] Wen Q, 2016, Teacher-Student Collaborative Assessment: A New Assessment Form Created by the Production-Oriented Approach. *Foreign Language World*, 5: 37–43.
- [29] Smale-Jacobse AE, Meijer A, Helms-Lorenz M, et al., 2019, Differentiated Instruction in Secondary Education: A Systematic Review of Research Evidence. *Frontiers in Psychology*, 10: Article 2366. <https://doi.org/10.3389/fpsyg.2019.02366>
- [30] Deunk MI, Smale-Jacobse AE, de Boer H, et al., 2018, Effective Differentiation Practices: A Systematic Review and Meta-Analysis of Studies on the Cognitive Effects of Differentiation Practices in Primary Education. *Educational Research Review*, 24: 31–54. <https://doi.org/10.1016/j.edurev.2018.02.002>
- [31] Parsons SA, Vaughn M, Scales RQ, et al., 2018, Teachers' Instructional Adaptations: A Research Synthesis. *Review of Educational Research*, 88(2): 205–242. <https://doi.org/10.3102/0034654317743198>
- [32] Pourdana N, Shahpouri Rad M, 2017, Differentiated Instructions: Implementing Tiered Listening Tasks in Mixed-Ability EFL Context. *Journal of Modern Research in English Language Studies*, 4(1): 69–87. <https://doi.org/10.30479/jmrels.2017.1566>
- [33] Rafi F, Pourdana N, 2023, E-Diagnostic Assessment of Collaborative and Individual Oral Tiered Task Performance in Differentiated Second Language Instruction Framework. *Language Testing in Asia*, 13: Article 6. <https://doi.org/10.1186/s40468-023-00223-7>
- [34] Sapan M, Mede E, 2022, The Effects of Differentiated Instruction (DI) on Achievement, Motivation, and Autonomy Among English Learners. *Iranian Journal of Language Teaching Research*, 10(1): 127–144. <https://doi.org/10.30466/ijltr.2022.121125>
- [35] Huang J, Mizumoto A, 2025, The Effects of Generative AI Usage in EFL Classrooms on the L2 Motivational Self System. *Education and Information Technologies*, 30: 6435–6454. <https://doi.org/10.1007/s10639-024-13071-6>
- [36] Chen A, Jia J, Li Y, et al., 2025, Investigating the Effect of Role-Play Activity with GenAI Agent on EFL Students' Speaking Performance. *Journal of Educational Computing Research*, 63(1): 99–125. <https://doi.org/10.1177/07356331241299058>
- [37] Guan L, Lee JCK, Zhang Y, et al., 2025, Investigating the Tripartite Interaction Among Teachers, Students, and Generative AI in EFL Education: A Mixed-Methods Study. *Computers and Education: Artificial Intelligence*, 8: Article 100384. <https://doi.org/10.1016/j.caeai.2025.100384>
- [38] Wu TT, Hapsari IP, Huang YM, 2025, Effects of Incorporating AI Chatbots into Think-Pair-Share Activities on EFL Speaking Anxiety, Language Enjoyment, and Speaking Performance. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2025.2478271>
- [39] Lamb M, 2017, The Motivational Dimension of Language Teaching. *Language Teaching*, 50(3): 301–346. <https://doi.org/10.1017/S0261444817000088>
- [40] Roseth CJ, Johnson DW, Johnson RT, 2008, Promoting Early Adolescents' Achievement and Peer Relationships: The Effects of Cooperative, Competitive, and Individualistic Goal Structures. *Psychological Bulletin*, 134(2): 223–246. <https://doi.org/10.1037/0033-2909.134.2.223>

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.