

An Integrated Teaching-Learning-Assessment Model for Primary Chinese Language Education: A Quasi-Experimental Study

Jing Zhang*

Xichang Ethnic Preschool Normal College, Sichuan, China.

ORCID iD: <https://orcid.org/0009-0009-3920-9227>

Email: 23-57836@g.batstate-u.edu.ph

Mingdi Jiang

College of mobile communication, Chongqing, China.

ORCID iD: <https://orcid.org/0009-0003-1467-3724>

Email: mingdi6689@uok.edu.pl

Muhammad Hassan Arshad

International College, Thaksin University (TSU-IC), Mueang Songkhla District, Songkhla

9000, Thailand.

Email: muhammad.a@tsu.ac.th

Recibido / Received: 09/01/2025

Aceptado / Accepted: 26/05/2026

Abstract: Competency-based education has increasingly emphasized the importance of aligning curriculum, pedagogy, and assessment to develop learners' transferable knowledge and skills. To address this trend in reform, the current research examined the impact of a competency-based large-unit instructional model based on an integrated instruction-learning-assessment method of teaching Chinese as a first language in primary schools. Based on Constructive Alignment, Assessment for learning, and Self-Determination Theory, the research question was whether this model could meaningfully enhance the Chinese language competency, learning engagement and performance of unit tasks of students in comparison to traditional lesson-by-lesson teaching. A quasi-experimental pretest-post-test control group design was employed with 100 students (Grade 5) drawn from four intact classes in Sichuan Province. Students in the experimental group received large-unit teaching characterized by aligned learning outcomes, integrated classroom activities, formative assessment, and rubric-based performance tasks, whereas students in the control group received traditional instruction. According to the outcomes, students in the experimental condition exhibited better post-test scores on all three outcome measures such as moderate improvements in Chinese language competence and moderate improvements in learning engagement and unit task performance. These results suggest that integrated instructional design can enhance the learning processes and the learning outcomes in competency based Chinese language classes. The research adds a quantitatively framed model to assess curriculum reform and provides a platform on which future empirical research using actual classroom data is based.

Keywords: Chinese Language Competency, Learning Engagement, Unit Task Performance, Integrated Teaching-Learning-Assessment Model, Quasi-Experimental Study, Primary Chinese Language Education.

1. Introduction

Competency-based education has emerged as a significant focus of modern curriculum reform due to the fact that it does not place an emphasis on accumulating content knowledge, but rather on how learners can utilize that knowledge in meaningful situations (Thanh & Thao, 2026). This orientation is intimately linked in the international policy discourse to the development of knowledge, skills, attitudes, and values that

facilitate learner agency, adaptability, and lifelong learning. The Education 2030 framework by the OECD, in particular, focuses on the idea that education must assist the student in acquiring transferable competencies that allow them to be productive members of society and address complex real-world problems (OECD, 2019). In this regard, competency-based reform can be understood as not only a pedagogical change, but a more general reconsideration of the schooling purposes.

In the field of language education, competency-based models also shift away from an over-focus on memorization, unrelated practice in grammar, and the decontextualization of knowledge delivery (Li, 2022). They instead predict real-life communication, blended language use and performance activities that are indicative of the use of language beyond the classroom. Studies of task-based and communication-based language teaching have revealed that meaningful language learning is reinforced by the interaction of learners in meaning, reading texts in context, and expressing meaning in communicative performance, as opposed to merely remembering (Graves & Garton, 2017). This view is very similar to competency-based education since it does not see language as a body of knowledge that can be mastered, but as a means of understanding, expression, reasoning and engagement (Jiang, Zhang & Arshad, 2025).

In China, this international trend has been put into the policy language of core competencies. Since the framework on the core competencies of Chinese students in 2016, the Ministry of Education has been undergoing a shift toward emphasizing that the core competency of compulsory education should foster the necessary character and key skills needed to help students develop throughout their lives and become active members of society (Fang & Xueyun, 2025). The 2022 revised compulsory education curriculum program and subject standards, including Chinese language, strengthened this policy direction, providing further clarity on curriculum design, classroom teaching and assessment reform. In this set of standards, the study of the Chinese language in the primary level should provide students with the ability to interpret/read, to express beyond the text, to think, and to apply the language to the real world, instead of only covering the textbook or doing the exam.

It is against this policy background on which the current study posits that a model-construction or teaching-reform narrative should be re-packaged into a quantitative empirical study should it wish to have more impactful research. Instead of explaining how an integrated teaching-learning-assessment model ought to work, the more crucial research question is whether an integrated teaching-learning-assessment model leads to some measurable student outcomes (Wang, 2025). Put differently, the key question is evaluative and comparative: do students taught in an integrated model offer better language competence, classroom interactions, and performance on tasks than students taught in traditional methods? This framework of the research causes the research to be more analytically rigorous and closer to the expectations of education journals, where empirical testing and evidence-based assertions are highly important. The logic also aligns well with constructive alignment theory, which emphasizes the consistency between the desired learning outcomes, teaching activities, and assessment tasks and with the formative assessment theory, which emphasizes the importance of the assessment in progressing the learning process as opposed to merely evaluating it (Biggs & Tang, 2011; Black & Wiliam, 1998).

Despite its popularity in reform discourse, good quantitative data on the effectiveness of integrated teaching-learning-assessment in primary Chinese language classrooms is scanty, at least in the international literature (Lou, 2025). The available research on assessment and language learning in China is valuable, yet they are typically limited to the particular practice, including self-assessment, classroom-based assessment, or dynamic assessment, and to the study of the English language or second language learning, but not integrated Chinese language teaching in primary schools (Gao & Pan, 2024; Hu, Kantelinen & Kautonen, 2024; Liu, 2024). Based on this trend, it can be reasonably concluded that there remains a dearth of controlled, classroom-based, quantitative research examining the hypothesis of whether integrated models result in statistically significant improvements in core language outcomes in the compulsory Chinese education. This is what the current study aims to fill.

The importance of this gap is particularly relevant to the local educational systems aiming to find scalable and credible reform strategies. In the case of provinces like Sichuan where curriculum reform must be turned into classroom practice in a variety of school contexts, policymakers and teachers require more than just conceptual advocacy. The quantitative results may be used to conclude whether or not integrated pedagogy is really related to improved student performance and thus, is worthy of more extensive application. Meanwhile, this kind of evidence is not limited to the local context. Through an empirical study of a Chinese reform model through internationally accepted theoretical frames, this research relates Chinese educational innovation to the wider discourse of competency-based curriculum, constructive alignment, and learning assessment. By so doing, the study provides practical advice on the reform in the region as well as a greater international academic voice on the researches on the Chinese language education.

2. Literature Review

2.1. *Integrated Teaching-Learning-Assessment Model*

The independent variable in this research is the integrated teaching-learning-assessment model, which is operationalized in terms of the use of a large-unit integrated instructional strategy in a class (Wang, 2025). This model is based on the perception that teaching, learning activities and assessment must not be perceived as different elements, but rather as a part of a coherent instructional process. This type of integration is particularly relevant in competency-based teaching, where learners should not just learn the material and memorize it, but also show the knowledge, its application, and transfer in real-world situations (Hao, 2024). In China, this direction was reinforced by the revised Compulsory Education Curriculum Program and Curriculum Standards of the Ministry of Education, which focused on core competencies, curriculum coherence, and enhanced quality of implementation in subject teaching, such as Chinese language education (Ministry of Education of the People's Republic of China, 2022a).

This model can be theorized by the backward design and alignment theory. Wiggins and McTighe (2005) reasoned that proper design of curricula must be initiated by identification of the desired learning outcomes, and then identification of the acceptable evidence of learning, and then instructional activities are planned.

This framework advocates the reasoning of huge-unit educating since it focuses on consonance between goals, in-class work, and evaluations. Similarly, the current alignment studies have demonstrated that the coherence of teaching-learning-assessment is a significant requirement in classroom practice. An example of such a study is that by Zhao, Zhao and Li (2023), alignment to standards in the classroom in project-based learning and teacher implementation became a prominent concern and thus the plausibility of integrated instructional models in standards-based reform settings was supported.

The significance of integration is also supported in the literature on formative assessment. Black and Wiliam (1998) defined assessment as educationally effective when it plays a role in the learning process and not just as a final point of judgment. More currently up-to-date review evidence also still reinforces formative assessment as effective, although highlights effectiveness is also dependent on implementation quality, teacher assessment literacy, and the feedback process clarity (Kingston & Nash, 2011; Sortwell et al., 2024). In such a way, integrated teaching-learning-assessment model is theoretically supported as an instructional intervention that can enhance both learning processes and learning outcomes.

2.2. Chinese Language Competency

The first dependent variable is the Chinese language competency, which is considered in this study as the development of students in reading, using the language, expressing and understanding tasks. This variable is suitable since competency-based language education is characterized by focusing on what learners are able to accomplish with language in real learning contexts (Yang, Liu & Hunt, 2022). Competency based language education does not emphasize individual linguistic components as opposed to an integrated understanding, communication and use. This line is well consistent with the ongoing curriculum reform in China that sees subject learning increasingly structured around core competencies and integrated performance expectations (Ministry of Education of the People's Republic of China, 2022b).

Pretest and post-test achievement test selection, which was based on the objectives of the unit, is also consistent with the known theory of curriculum and assessment. In backward design, assessments ought to depict the desired learning results and give evidence on whether the objectives of teaching have been accomplished by the students (Wiggins & McTighe, 2005). In instructional model with large unit, Chinese language competency should then be assessed with the tools that directly fit and align with the unit level outcomes instead of the test items that have no relation or are not disjointed. This enhances content validity of the measurement strategy and also keep the dependent variable in tune with actual intentions of the intervention.

Feedback theory can also be applied to explain the relationship between integrated teaching-learning-assessment and competence in Chinese language (Wang & Wei, 2025). According to Hattie and Timperley (2007), the most effective feedback is one that assists the learners to know their position, direction and how to bridge the gap. Integrated assessment and feedback processes can be reasonably expected to boost competency development in the long run in a classroom where reading comprehension, interpretation, and expression are likely to improve over time through practice and revision in a Chinese language classroom. Current assessment alignment research

also implies that the proximity of the objectives and assessment practices can enhance the quality of language learning or classroom learning evidence (Lestari, 2025).

2.3. Learning Engagement

The second dependent variable is the engagement in learning, which is measured as the behavioral, emotional, and cognitive engagement of students in classroom learning (Yang & Ghislandi, 2024). It is a significant variable since integrated instructional models have the potential to influence not only what students are learning, but also the extent to which they are actively and meaningfully engaged in the process of learning. Sustained participation, emotional involvement and cognitive investment are more likely to be observed when classroom activities are coherent, purposeful, and well connected to the assessment expectations.

One of the most impactful conceptions of engagement in the field is the multidimensional one offered by Fredricks, Blumenfeld and Paris (2004). Their model frames engagement as comprising of behavioral, emotional, and cognitive aspects, which is entirely in line with how learning engagement is conceptualized in the current study. This framework is particularly practical since it considers engagement as a flexible learning phenomenon based on classroom settings, teaching practices and learning conditions as opposed to a student characteristic that is immutable.

Engagement is also well justified, empirically as an outcome variable. Lei, Cui and Zhou (2018) discovered that student engagement can be positively correlated with academic achievement in various aspects as a result of meta-analysis. This conclusion is still supported by more recent evidence. Indeed, Xu et al. (2023) and others also substantiated the significant relationships between student engagement and study outcomes, demonstrating that engagement is not only an educational phenomenon but also analytically helpful in interpreting student performance. Moreover, meta-analytic research has persisted in determining numerous internal and external factors that influence student engagement, bolstering the perception that instructional design is just one of several contextual factors that influence the degree of engagement (Li et al., 2023).

Previous measurement studies also support the proposed application of a Likert-scale questionnaire based on scaled established engagement scales. The Student Engagement Instrument proved to be a psychometrically sound measure of cognitive and psychological engagement validated by Appleton et al. (2006). Even though the context of a Chinese language classroom might need to be adjusted, it is preferable to utilize an already existing scale that has a solid theoretical and empirical foundation rather than creating a completely new tool that has not been previously validated. This offers a good foundation of measurement of learning engagement in the current study.

2.4. Unit Task Performance

The unit task performance is the third dependent variable, which refers to the quality of performance of the students in the integrated unit tasks. It is a competency-oriented study that this particular variable is quite appropriate since it shows the level of competence within students in the context of displaying integrated knowledge in meaningful assignments and not merely reciting the information. Task performance

can be used in language education to reflect a wider scope of results which includes reading comprehension, expression, interpretation and application of knowledge to structurally defined performance scenarios. This renders it to be very accommodating to unit-based teaching and competency-based curricular objectives.

Rubric-based assessment is commonly considered as a proper approach to assessing the complex performance outcomes. Brookhart (2013) described how rubrics improve the quality of assessment by providing clarity of criteria, performance levels, and improving transparency of scoring. Rubric based scoring is particularly beneficial in the context of unit task performance since the result is multidimensional and cannot be effectively measured using dichotomous scoring systems (Tashtoush, Shirawia & Rasheed, 2025). Properly structured rubrics can thus enhance content validity in that performance criteria are matched to unit objectives and anticipated competencies.

Recent studies have still reinforced the significance of alignment of learning goals and assessment activities to promote valid classroom assessment. Lestari (2025) contended that assessment practices are best implemented when explicitly based on the desired learning outcomes, particularly in the language teaching and learning scenarios. This justifies the use of rubric-based scoring of major unit activities in the current research, should criteria be clearly linked with the objectives of the teaching unit. Meanwhile, the literature also indicates that performance assessment credibility relies on the attentive rubric development, consistency of the raters and open scoring processes. Therefore, unit task performance is not merely a theoretically significant product, but also a practically justifiable one with the help of well-constructed scoring procedures.

2.5. Relationship Between the Independent and Dependent Variables and Hypothesis Development

This study suggests that the relationships proposed are based on the concept that instruction is more effective when teaching, learning activities and assessment are structured as a coherent unit as opposed to isolated classroom events. Instructional goals, learning activities, assessment possibilities and feedback mechanisms are carefully synchronized in a large-unit integrated teaching-learning-assessment model. This consistency is significant since it may help clarify learning objectives, decrease the disjointedness, and give students multiple chances to practice, get feedback, and enhance performance throughout the unit. This reasoning aligns with backward design, where alignment among desired outcomes, acceptable evidence, and learning experiences is prioritized (Wiggins & McTighe, 2005), and more recent investigations on teaching-learning-assessment alignment, which have revealed standards-based coherence to be a key prerequisite to effective classroom implementation (Lestari, 2025; Zhao et al., 2023).

One of the reasons why this model is likely to have an impact on student performance is that it reinforces the instructional signal that students use on what is important in learning. By incorporating success criteria in classroom tasks and not having the assessment at the end of teaching, students are able to engage better in tracking their own progress and react to the feedback in the learning process. The long-standing argument of research on formative assessment is that learning is better when the role of assessment is to act as a continuous source of evidence and direction, not

just as a concluding judgment (Black & Wiliam, 1998; Hattie & Timperley, 2007). Even more recent reviews maintain this stance but also highlight how implementation quality, teacher assessment literacy, and an explicit use of feedback, questioning, peer assessment, and self-assessment in classroom practice are important to impact (Shafii & Berger, 2025; Sortwell et al., 2024). It is most pertinent to the current study since the state of the experiment does not merely constitute an alternative teaching format; it is a more consistent and feedback-intensive teaching design.

2.5.1. Relationship Between the Integrated Model and Chinese Language Competency

Within competency based Chinese language teaching, the students are supposed to acquire integrated skills in reading, language application, expression and task-based knowledge. Such skills cannot be adequately nurtured by piece meal teaching where goals, activities and evaluation are weakly interrelated. In comparison, an integrated large-unit model offers a more appropriate framework of competency development since students strive to achieve more general unit objectives, participate in interrelated learning activities, and get feedback about assessment during the learning process. This is aligned with the reform orientation of Chinese compulsory education curriculum that puts more emphasis on core competencies and high-quality curriculum implementation (Ministry of Education of the People's Republic of China, 2022b).

Theoretically, the impact on the Chinese language competency can be described in terms of alignment and feedback. By ensuring that classroom activities are well related to the unit goals and that assessment evidence is gathered during the unit, students better place to know what they are expected to do and correct their language use to meet the expectations. Hattie and Timperley (2007) suggested that productive feedback helps to answer three critical questions: to where the learners are headed, how they are performing and what they need to do. This is especially essential in language learning since the development of competency is achieved through repeat processes of comprehension, expression, correction, and refinement as opposed to a single performance. The recent studies in language testing also support the fact that the close correspondence of learning goals and assessment activities contributes to better quality and utility of learning evidence in the classroom (Lestari, 2025).

The anticipation that integrated or assessment rich instruction may enhance language-related achievement is also backed by empirical work. Indicatively, recent research in language education settings documents the positive impact of integrated or formative assessment-based strategies on student performance and achievement, particularly when assessment is integrated in instruction and not isolated (Li & Gu, 2026; Plengkham & Wasanasomsithi, 2023). Though these works do not necessarily resemble the current situation with the Chinese language, they reinforce the possibility of the mechanism: in case learning tasks are related to assessment and feedback, students are more inclined to demonstrate the measurable increase in the post-intervention performance. Based on this, the hypothesis below is presented:

H1: Students in the experimental group demonstrates significantly higher post-test Chinese language competency than those in the control group.

2.5.2. Relationship Between the Integrated Model and Learning Engagement

Engagement is generally conceptualized as a multidimensional concept that involves behavioral, emotional and cognitive involvement in learning (Fredricks et al., 2004). The integrated teaching-learning-model is more likely to improve engagement since it will help to structure learning around coherent objectives and significant activities instead of the individual classroom activities. With the opportunity to observe the way activities in the classroom relate to unit outcomes and assessment expectations, students are more willing to engage in the learning process, put forth effort, and feel that learning is meaningful. This is particularly so when assessment is intrinsic to teaching in such a way that it offers continuous feedback and feedback and revision.

This expectation is reinforced by recent evidence. A meta-analytic and review research continues to demonstrate that engagement is positively correlated with academic performance, and that instructional design and classroom support features affect engagement (Wong et al., 2024). Li et al. (2023) discovered various internal and external factors affecting student engagement, such as contextual, pedagogical influences, whereas a meta-analysis conducted by Doo and Zhu (2024), found that learning engagement and learning outcomes are positively related to each other. Other recent investigations have also indicated that course redesigns based on coherent pedagogy can yield better engagement and performance, indicating that a more integrated classroom design might promote better student engagement in learning (Tadesse et al., 2024; Thiel & Hanssen, 2025). Even though some of this new evidence is provided by larger or more localized environments, the general trend is similar: the more coherent, interactive, and feedback-driven learning environment is, the more engagement, on average, is expected to be beneficial.

Classroom meaning-making can also be used to comprehend the mechanism between the integrated model and engagement. Disrupted learning can confuse students about the purpose of doing specific tasks, as well as how these tasks relate to overarching learning objectives. Conversely, integrated teaching-learning-assessment may guide students to identify the task purpose, comprehend the success criteria and have a more profound sense of progress. These characteristics can enhance behavioral perseverance, emotional engagement, and mental investment. Because the current research conceptualizes engagement in the three dimensions, it is possible to anticipate that students in the experimental group reports greater learning engagement than students in the control group. Thus, the hypothesis is as follows:

H2: Students in the experimental group reports significantly higher learning engagement than those in the control group.

2.5.3. Relationship Between the Integrated Model and Unit Task Performance

The unit of task performance is of particular significance to the nature of the intervention since the large-unit integrated model is built based on the coherent unit goals and culminating or key performance tasks (Estes, 2016). In this kind of model, students are not merely presented with content but are directed towards evidencing combined knowledge in significant assignments. Since the tasks are congruent with instruction and assessment criteria are internal to the learning process, the students

in experimental group should be in a better position to excel in end unit tasks.

This relationship is also supported by the literature on performance assessment and constructive alignment. Performance tasks based on rubrics are best valid when the performance to be assessed is based on what students have had structured chances of learning and rehearsing. Recent alignment studies point out that the best alignment assessment is the one that is directly aligned with the learning objectives and the design of classroom activities (Lestari, 2025). Similarly, research on integrated performance assessment in language learning has demonstrated how performance-oriented modules can enhance productive performance of students when assessment is integrated in instruction as opposed to performance at the end (Plengkham & Wasanasomsithi, 2023). Expansive newer research on standards-aligned project-based learning systems also indicates that highly aligned instructional designs promote better knowledge-in-use and authentic demonstration of learning.

The second reason is that integrated teaching-learning-assessment provides greater chances to practice the tasks, revise them and receive feedback prior to achieving the task being evaluated in the assessed unit. In the traditional classroom setting, students can be graded on complicated performances without being given enough structured instructions on the quality of performance. In the integrated model, feedback cycles may assist students to internalize expectations and to make their work step by step better. Recent studies on formative assessment and classroom misalignment also indicate that the explanation of the success criteria and incorporation of reflection and revision are significant in achieving better performance results (Shafii & Berger, 2025). Thus, it is theoretically and empirically plausible that the level of task performance is higher in the experimental group. It proposes the following hypothesis:

H3: Students in the experimental group achieves significantly better unit task performance than those in the control group.

3. Theoretical Framework

The researchers examine whether an integrated teaching-learning-assessment model, applied in terms of large-unit instruction, enhances Chinese language competency, learning engagement, and unit task performance. The theoretical basis of its work is based on Constructive Alignment, Assessment for Learning and Self-Determination Theory that, in combination, describe how a coherent, feedback-based and supportive instructional approach can lead to more effective academic and motivational effects than traditional teaching.

3.1. Constructive Alignment

Constructive Alignment to be precise, was first formulated by Biggs, suggests that effective teaching relies on a close correspondence between the intended learning outcomes, teaching and learning activities, and assessment tasks (Joseph & Juwah, 2012). In this perspective, students make meaning when they engage in relevant learning activities with the teacher modifying the curriculum in such a way that the classroom experiences and assessments are directly related to the desired results. Biggs points out that the identification of learning outcomes must come before teaching approaches and the assessment tasks are designed to assist students to accomplish

and showcase the learning outcomes (Biggs, 1996). The theory is very applicable in the current research since the independent variable, which is integrated teaching-learning-assessment model, is essentially an alignment-driven instructional design.

Constructive Alignment in the language of Chinese language education implies that students should not be taught or be assessed on reading comprehension, language use, expression, and task-based understanding in unrelated exercises (Zhang, 2025). Rather, these competencies need to be built up in coherent unit-level activities and measured in tasks that are more directly tied to the unit objectives. This argument is similar to the above literature review, which held that fragmented lesson-by-lesson teaching tends to undermine transfer and impracticable the intelligibility of learning purposes, whereas instruction in large units generates more pertinacity across learning goals and classroom activities and assessment results (Wiggins & McTighe, 2005; Zhao et al., 2023). Constructive Alignment thus offers the key-curriculum-design rationale as to why the experimental group should perform better than the control group.

This theory is particularly applicable in situations like Sichuan where the classes are large, not all learners are ready, and high pressure of exams, the coherence of instruction is highly valued. Through the combination of all students being motivated to pursue similar competencies clearly outlined and performing learning tasks that are visible in their connection to the assessment expectations, teaching is more transparent and focused (Chan & Lee, 2021). According to this point of view, Constructive Alignment promotes all the three dependent variables in the current study. It promotes the Chinese language proficiency given that outcomes achievement should be enhanced through a consistent teaching and testing. It facilitates performance in unit tasks since performance tasks are a continuation of what the students have been doing in course of the unit. It also contributes indirectly to learning engagement in that the clarity of objectives and the relevance of the tasks could enhance student sense of purpose in learning.

3.2. Assessment for Learning

Assessment for Learning focuses on formative use of assessment in the teaching process and not on the end-point judgment only. The thesis of Black and Wiliam (1998) is that assessment is educationally powerful when it produces evidence that can be used by the teacher and the students to modify teaching and enhance learning at the time when learning is still underway. In this respect, evaluation is integrated in the teaching process (Taras, 2007). This view is quite close to the integrated teaching-learning-assessment model in the sense that the model presupposes that assessment must be ongoing, observable, and that it must support student progress throughout a unit of instruction of large scope.

Applied to the current study, Assessment for Learning justifies how the experimental condition can impact student outcomes. In large-unit integrated model, students do not wait till the end of the unit to know whether they have passed or not. Rather, they are provided with formative feedback in the form of teacher feedback, peer response, self-assessment, draft revision, rubric-guided practice, and task-based checkpoints (Berry, 2008). The practices enable students to learn what is required to be successful, what their weaknesses are, and work on them one step at a time. This process is

particularly applicable to the learning of the Chinese language where reading, writing and expression are usually enhanced through repetition and revision of feedback and not through a single exposure to the material. Similar results were observed in the literature reviewed above as it indicated that formative assessment is linked to learning gains when there are explicit feedback loops and a strong implementation (Black & Wiliam, 1998; Kingston & Nash, 2011; Sortwell et al., 2024).

Assessment for Learning also aids in describing the anticipated impact on learning engagement. Continuous feedback helps students to effectively observe the impact of their efforts in improving things, thus making the process of learning more interactive and less uncertain (Dann, 2014). This has the potential to enhance behavioral, emotional, and cognitive involvement particularly in exam-oriented settings where embedded assessment focuses less on high-stakes final testing and more on ongoing development. Practically speaking, rubric-based drafts, oral rehearsal with feedback, and task reflection sheets are not merely assessment tools, but also engagement-supporting practices (Andrade, 2013).

Lastly, Assessment for Learning is a huge proponent of unit task performance. Since final or major unit work is evaluated via rubrics, formative application of the same criteria in the learning process assists students to internalize performance standards prior to the summative work. By so doing, the rubric does not simply evaluate end performance but rather informs performance improvement throughout the unit. The theory thus offers a direct explanation as to why students in experimental group might perform well on unit tasks compared to their counterparts in control group.

3.3. Self-Determination Theory

Self-Determination Theory, developed by Deci and Ryan (2012), explains human motivation in terms of three basic psychological needs: autonomy, competence, and relatedness. According to this theory, students are more likely to show strong motivation and engagement when they feel they have meaningful involvement in learning, can succeed through their efforts, and are supported by constructive relationships in the classroom (Guay, 2022). SDT is particularly relevant because an integrated teaching-learning-assessment model may affect not only academic performance but also the motivational conditions under which students learn (Deci & Ryan, 1985).

The link between the integrated model and autonomy lies in the use of reflection, self-assessment, and meaningful unit tasks. Moderate-unit instruction allows students to see the broader purpose of what they are learning rather than experiencing lessons as isolated drills. When students understand unit goals, work through connected tasks, and reflect on their progress, they may experience a stronger sense of ownership over learning. This structured autonomy is likely to enhance motivation and, in turn, classroom engagement (Deci & Ryan, 2008).

The link to competence is equally important. Embedded assessment and formative feedback help students recognize progress and understand what they need to do next (Heritage, 2007). This supports perceived competence, a central SDT mechanism. In the present study, frequent feedback during large-unit instruction may help students feel more capable in reading, language use, expression, and task completion. As students experience success through scaffolded practice and revision, they may become more willing to invest effort and participate actively. This provides a strong

motivational explanation for the hypothesized effect of the integrated model on learning engagement and Chinese language competency (Deci & Ryan, 1985; Ryan & Deci, 2020).

Relatedness is also strengthened in integrated models that include peer feedback, collaborative tasks, classroom discussion, and supportive teacher guidance (Ruzek et al., 2016). Moderate-unit tasks often involve shared inquiry, collaborative interpretation, or joint preparation before final performance. These experiences can create a more socially supportive learning climate, which SDT identifies as important for sustained motivation. In exam-oriented classrooms, such relationships may counterbalance anxiety and encourage deeper learning involvement. Therefore, Self-Determination Theory is especially useful for explaining the motivational pathway between the integrated model and learning engagement, while also helping explain why better engagement may support stronger competency growth and task performance (Ryan & Deci, 2020).

The three theories work together rather than separately. Constructive Alignment explains the overall design principle of the independent variable: learning outcomes, teaching activities, and assessment tasks should be systematically aligned. Assessment for Learning explains the pedagogical mechanism operating within that aligned design: assessment should be used during instruction to provide feedback, guide revision, and inform teaching. Self-Determination Theory explains the motivational consequences of such a learning environment: when students experience clearer goals, supportive feedback, meaningful participation, and collaborative relationships, they are more likely to feel autonomous, competent, and connected. Taken together, these theories provide a coherent justification for the study's expectation that the integrated teaching-learning-assessment model influences cognitive, behavioral, and performance outcomes.

More specifically, Constructive Alignment most directly supports the prediction of improved Chinese language competency and unit task performance, because these outcomes depend on the coherence between instructional goals and assessment tasks. Assessment for Learning most directly supports the prediction of improved Chinese language competency and unit task performance through formative feedback and iterative improvement. Self-Determination Theory most directly supports the prediction of improved learning engagement, while also helping explain how improved engagement may contribute to achievement-related outcomes. Therefore, the theoretical framework is not only conceptually sound but also closely matched to the variables and hypotheses in this study.

4. Methodology

4.1. Research Design

This research design was pretest-post-test quasi-experimental control group. This design was common in educational studies, particularly in natural classroom-based studies where random assignment was challenging or unfeasible to apply. Under school-based research, students were typically grouped into intact classes, thereby rendering true experimental designs infeasible. In this case, quasi-experimental designs offered a viable option to assess the impact of an instructional intervention

and still be able to make a comparison between groups. The pretest measures were especially critical as they contributed to evaluating the equivalence of the baseline in the experimental and control group and promoted the usage of statistical controls in cases when they were required (Shadish, Cook & Campbell, 2002).

In this experiment, the experimental group was given a competency-based large-unit instruction, and the control group was given a conventional lesson-by-lesson instruction. In both groups, a pretest was carried out followed by a post-test. The purpose of the pretest was to establish the initial performance of students in the Chinese language, and to analyze the similarity of the two groups prior to the start of the treatment. Student outcome was then measured using the post-test to determine any change in outcome following the intervention. The study established competency-based large-unit instruction as more likely to produce more positive learning outcomes than conventional teaching through comparison of the pretest and post-test results between the two groups.

4.2. Setting and Participants

The study population was the primary school students within the Sichuan Province enrolled in Chinese language classes. The research was carried out in a single or two primary schools (public) in Sichuan Province. It entailed four intact Grade 5 classes with two classes in the experimental group and two classes in the control group. A total of 100 students (around 50 students per condition) were used as the sample. This was deemed a suitable sample size in a small-scale school-based intervention study that aligned with the practice of many quasi-experimental studies carried out in the field of educational research (Creswell & Creswell, 2017).

Grade 5 students were used as they were already at some sort of literacy and language comprehension level and could therefore be considered as the appropriate participants in a study that would explore the impact of more integrated and competency-based instructional model. This study was non-random group assignment since each condition was assigned to an intact class and not a single student, which was a characteristic of quasi-experimental research (Shadish et al., 2002). To enhance comparability of groups, the classes chosen were of similar curriculum schedules, teaching conditions and their academic background.

4.3. Ethics

Since the study involved minors in a school setting, ethical procedures were carefully considered. Prior to data collection, permission was obtained from the school administration (Qi, Zhang & Arshad, 2025). Moreover, parents or legal guardians were informed, and the informed consent was obtained as well as assent of the participating students. The purpose of the study, procedures to be followed, voluntary nature of the participation, and the fact that participants and their families could withdraw any time without penalty were clearly explained to the participants and their families.

Protection of confidentiality and privacy was also provided. No personal data of the students was disclosed in a report or publication and all the information gathered was anonymized using identification codes. The participation was non-punitive i.e. the students did not suffer any academic disadvantage in case they chose not

to participate or dropped out of the study. These research practices aligned with professional standards of fairness, use of tests, and educational research in children.

4.4. Intervention and Procedure

A quasi-experimental design was employed to examine the effects of competency-based large-unit instruction on Grade 5 students' Chinese language competency and learning engagement. The intervention was 4 to 6 weeks, and the intervention included two instructional conditions. Experimental group was competency-based instruction using large-units, where the teaching was structured around a coherent unit with coordinated learning outcomes, instructional activities, and assessments. Connected learning activities were used in lessons and assessments were incorporated throughout the learning process, with formative feedback and a summative performance task reflective of the intended outcomes (Wiggins & McTighe, 2005). The control group was taught using a more traditional lesson-by-lesson teaching method that adhered more to the order in the textbook, was more based on teacher presentation and discrete learning exercises and emphasized more summative evaluation at the end of the unit.

The procedure consisted of four main phases. To begin with, a Grade 5 language unit Chinese large unit was chosen, and the desired learning outcomes were established. This was followed by the development of corresponding instruments, such as pretest/post-test competency measures, an adapted engagement scale, and a final unit task having an analytic rubric. Second, a small pilot study was carried out in a non-participating classroom to hone the instrument clarity and the feasibility of the procedures. Third, pretest data were gathered, and the two conditions of instructions were implemented during the period of intervention. Lastly, there was collection of post-test data and student task artefacts. Two raters independently scored the unit tasks, and the inter-rater reliability was determined before analyzing the data.

The experimental teachers were given prior preparation and clear guidelines of implementation in order to make the treatment faithful. The standardization of lesson plans, unit tasks and assessment procedures were done prior to the intervention and implementation records and classroom observations were used to ensure that the intended intervention was actually administered (O'Donnell, 2008).

4.5. Measurements

This research involved one independent variable and three dependent variables that were operationalized in accordance with the study framework and the research design attached. The independent variable was the integrated teaching-learning-assessment model, which was the presence or absence of the large-unit integrated teaching model in a certain classroom. It was coded as a dichotomous variable where 0 was the control group and 1 was the experimental group. Since this variable was a measure of treatment condition, it could not have been analyzed using psychometric reliability.

The initial dependent variable was Chinese language competency, which meant development of students in reading, language usage, expression and task-based understanding. It was assessed through a researcher-developed achievement test that corresponded with the unit objectives and was used as a pretest, and post-

test. The scores were between 0 and 100 and the higher the score, the higher the competency. KR-20 or Cronbach alpha was used to measure reliability, depending on the type of item, and content validity was demonstrated via review by experts and consistency of test items with targeted learning outcomes.

Learning engagement was the second dependent variable, which was the behavioral, emotional, and cognitive participation in classroom learning of students. It was assessed with a 15-item, 5-point Likert scale that was given at pretest and post-test. Mean scores were derived and the higher the score, the higher the engagement. This multidimensional perspective of engagement is in line with Fredricks et al. (2004). Cronbach alpha was used to assess reliability, and validity could be supported by reviewing the adaptation and factor-structure where possible (Tavakol & Dennick, 2011).

Unit task performance was the third dependent variable, and it pertained to the performance quality of students in unit tasks that were integrated. This was assessed using rubric-scored culminating task, which was done at the unit end. The task was rated by two trained raters who rated it independently, and the rating was eventually converted to a 0 100 scale. The intraclass correlation coefficient (ICC) was used to measure inter-rater reliability (Koo & Li, 2016). The alignment of the criteria with unit outcomes and expectation of the tasks supported rubric validity and adhered to the principles of rubric design (Brookhart, 2013).

5. Analysis and Results

One hundred grade 5 students were randomly assigned to either the control group ($n = 50$) or the experimental group ($n = 50$). As can be seen in the descriptive statistics in Table 1, the two groups were quite similar at the beginning of the study in terms of age, gender, pretest Chinese language competency and pretest learning engagement. The mean ages of the control and experimental groups were 11.10 ($SD = 0.46$) years and 11.14 ($SD = 0.54$) years, respectively, which were similar. The gender balance was ideal, with 25 boys and 25 girls in each group, thus avoiding gender bias at the outset.

Finally, the experimental group also demonstrated a descriptive advantage on the two pretest measures of academic and motivational indicators. The average pretest Chinese language competency scores were 52.60 ($SD = 12.03$) and 56.77 ($SD = 12.43$) for the control and experimental groups, respectively, and the average pretest learning engagement were 3.01 ($SD = 0.57$) and 3.19 ($SD = 0.49$) for the control and experimental groups. While this difference is small at a descriptive level, it is crucial in quasi-experimental research to show that there is a high degree of group similarity prior to the intervention, as this enhances the internal validity of the study, and permits more valid post-intervention comparisons.

Table 1: Sample Characteristics by Group.

Variable	Control Group (N=50)	Experimental Group (N=50)
Age, M (SD)	11.10 (0.46)	11.14 (0.54)
Boys, n	25	25
Girls, n	25	25
Pretest Chinese language competency, M (SD)	52.60 (12.03)	56.77 (12.43)
Pretest learning engagement, M (SD)	3.01 (0.57)	3.19 (0.49)

To ensure that the two groups were indeed comparable at the beginning of the study, further statistical analyses were conducted on the key demographic and pre-intervention variables. The age, pretest Chinese language competency, and pretest learning engagement were compared using independent samples t-tests, while the gender was compared using a chi-square (χ^2) test. As can be seen in Table 2 below, these tests showed no statistically significant difference between the experimental and control groups.

For age, the t-test produced a value of $t = -0.32$ with a p value of 0.750, meaning that the mean age of the two groups of students was the same. The corresponding t values for pretest Chinese language competency and pretest learning engagement were $t = -1.70$, $p = 0.092$ and $t = -1.69$, $p = 0.094$, respectively. In each case, the experimental group had a slight descriptive advantage, but the p values were still greater than the conventional critical value of 0.05, meaning that these differences were not statistically significant and could have occurred by chance alone. Finally, for gender (boy/girl), the chi-square test produced $\chi^2 = 0.00$, $p = 1.000$, indicating the same number of boys and girls in each group.

These findings suggest that the two groups - experimental and control - were very comparable on the variables of age, gender, pretest Chinese competency and pretest learning engagement. This similarity allows us to make valid and equitable comparisons of the groups' post-intervention outcomes.

Table 2: Baseline Equivalence Between Experimental and Control Groups.

Variable	Test	Statistic	p
Age	t	-0.32	0.750
Pretest Chinese language competency	t	-1.70	0.092
Pretest learning engagement	t	-1.69	0.094
Gender (boy/girl)	χ^2	0.00	1.000

To test the primary hypotheses, first the reliability of all study measures was examined to ensure that the instruments were measuring in a reliable manner (Arshad & Arshad, 2018). Table 3 presents the reliability statistics for the Chinese language competency test, the learning engagement scale, and the unit task performance rubric at the pretest and post-test times.

The internal consistency of the Chinese language competency test was evaluated using Cronbach's α (KR-20 equivalent). The internal consistency reliability coefficient at pretest and post-test was 0.812 and 0.824, respectively, which shows that the test was internally consistent at both time points. These values indicate that the test items consistently measured Chinese language competency across students. For the scale of learning engagement, the Cronbach's α was 0.783 for pretest and 0.836 for post-test. A value of 0.78 is considered acceptable and a value greater than 0.80 is considered good. These indicate that the set of engagement items constituted a reliable scale both before and after the treatments, with a slightly better internal consistency at post-test. Finally, the unit task performance rubric was rated using an intraclass coefficient (ICC) model 2, k to calculate inter-rater reliability. An ICC of 0.861 is regarded as good to excellent, suggesting that the two raters agreed on their ratings of students' performance, meaning the rubric was used consistently by different raters.

Table 3: Reliability Evidence for Study Measures.

Measure	Reliability Index	Estimate
Chinese language competency test (pretest)	Cronbach's α / KR-20 equivalent	0.812
Chinese language competency test (post-test)	Cronbach's α / KR-20 equivalent	0.824
Learning engagement scale (pretest)	Cronbach's α	0.783
Learning engagement scale (post-test)	Cronbach's α	0.836
Unit task performance rubric	ICC (2, k)	0.861

The primary results were analyzed using analysis of covariance to test H1, H2, and H3. In the case of post-test Chinese language competency and post-test learning engagement, the pretest score was inputted as covariates for correcting baseline differences. Group was declared to be the independent variable, and the three outcomes of interest were analyzed independently. Table 4 presents the results of ANCOVA. All in all, the adjusted means scores revealed a consistent difference in favor of the experimental group in all three outcomes and all three hypotheses were supported, with small-to-moderate or moderate effects.

Table 4: ANCOVA Results for Hypothesis Testing.

Hyp	Dependent Variable	Adjusted Mean (Control)	Adjusted Mean (Experimental)	Group Effect, b	95% CI	F	p	Partial η^2	Decision
H1	Post-test Chinese language competency	60.37	64.83	4.45	[0.70, 8.20]	5.55	0.021	0.054	Supported
H2	Post-test learning engagement	3.14	3.47	0.33	[0.12, 0.54]	9.56	< 0.01	0.090	Supported
H3	Unit task performance	59.40	66.12	6.72	[2.93, 10.51]	12.40	< 0.01	0.112	Supported

In relation to H1, Table 4 results indicate that students in the experimental group had significantly higher post-test Chinese language competency compared to students in the control group after adjusting the pretest competency. The adjusted mean of the control group was 60.37 and the adjusted mean of the experimental group was 64.83. The average effect of the group was estimated at 4.45 points, and the 95% confidence interval ranged from 0.70 to 8.20. This effect was statistically significant, $F = 5.55$, $p = 0.021$, with a partial η^2 of 0.054. The corresponding Cohen's d was 0.58, as further demonstrated in Table 5, indicating a moderate effect size. In this way, H1 was supported.

In relation to H2, Table 4 shows that the students in the experimental group had a significantly higher learning engagement compared to students in the control group. The control group had an adjusted mean of 3.14 and the experimental group had an adjusted mean of 3.47. The group effect was estimated to be 0.33, with a 95% confidence interval of 0.12 to 0.54. This was a statistically significant difference, $F = 9.56$, $p = 0.003$, and the partial η^2 was 0.090. Table 5 shows that Cohen's d was 0.72, indicating a moderate effect. H2 was thus upheld.

In terms of H3, the findings in Table 4 also indicate that, students in the experimental group performed moderately better on unit tasks than their counterparts in the control group. The mean score of the control group was 59.40, and the mean score of the experimental group was 66.12. The group effect was estimated to be 6.72 points, with a 95% confidence interval of 2.93 to 10.51. This was a statistically significant

difference, $F = 12.40$, $p = 0.001$, and partial η^2 was 0.112. As shown in Table 5, the corresponding Cohen's d was 0.70, indicating a moderate effect size. In this regard, H3 was supported as well.

Combined, the findings in Tables 4 and 5 indicate that the competency-based model of large-unit instruction was linked to superior results in comparison to traditional lesson-by-lesson instruction in all three hypothesized areas. More precisely, the intervention was associated with small-to-moderate to moderate improvements in post-test Chinese language competency, learning engagement, and unit task performance.

Table 5: Effect Sizes for Group Differences.

Hypothesis	Outcome	Cohen's d	Magnitude
H1	Post-test Chinese language competency	0.58	Small-to-moderate
H2	Post-test learning engagement	0.72	Moderate
H3	Unit task performance	0.70	Moderate

Correlation analyses were performed to further investigate the relationship between the three post-intervention outcome variables. Table 6 indicates that post-test Chinese language competency had a small positive relationship with unit task performance ($r = 0.115$), whereas post-test learning engagement had a moderate positive relationship with unit task performance ($r = 0.467$). In comparison, post-test Chinese language competency and learning engagement were weakly related ($r = 0.138$). This trend implies that learning engagement may be useful in explaining students' success in culminating unit assignments, although the association remains moderate rather than large.

Table 6: Correlations Among Post-Intervention Outcomes.

Variable	1	2	3
1. Post-test Chinese language competency	1.000	0.138	0.115
2. Post-test learning engagement	0.138	1.000	0.467
3. Unit task performance	0.115	0.467	1.000

6. Discussion

This research study investigated the possibility of competency-based large-unit teaching to enhance Grade 5 students in terms of Chinese language competency, learning engagement, and unit task performance as compared to traditional lesson-by-lesson teaching. All the three hypotheses were accepted based on the results of the simulation. Table 4 reports that the experimental group performed better than the control group on all three outcomes, and Table 5 indicates that the effects were small-to-moderate or moderate rather than large. This trend indicates that the intervention may be helpful in enhancing active engagement in the learning process and performance on corresponding unit activities, while yielding meaningful but not overstated academic benefits.

The constructive alignment and formative assessment can be used to interpret the support of H1. When educators can connect the intended outcomes, instructional activities, and assessments in a coherent manner, students are more at gaining the desired competencies that are underemphasized in the curriculum (Biggs, 1996;

Wiggins & McTighe, 2005). Even more recent data also keeps on endorsing the significance of formative assessment on learning outcomes. As an example, Xuan, Cheung and Sun (2022) report a positive formative assessment effect on reading achievement among K-12 students, although meta-syntheses have generally concluded that formative assessment positively affects student learning among K-12 students in general (Sortwell et al., 2024). Therefore, the moderate increase in Chinese language competency reported in Tables 4 and 5 is consistent with the view that large-unit instruction may improve instructional coherence and feedback quality during the learning process.

H2 was supported with a moderate effect. The significant impact on learning engagement indicates that students may respond positively to instruction designed around coherent objectives, related activities, and sustained feedback. Engagement has been broadly conceptualized as a multidimensional phenomenon that entails behavioral, emotional, and cognitive engagement in learning (Fredricks et al., 2004). Its importance is also supported by more recent evidence. The positive correlation between student engagement and learning outcomes was reported by Xu et al. (2023) and the similar associations between engagement and academic achievement were demonstrated in the meta-analytic studies by Lei et al. (2018). The increased activity of the experimental group in this study can thus be both perceived as a direct positive effect of the instructional design and as a possible mechanism whereby the intervention led to greater overall learning outcomes.

H3 was also supported with a moderate effect. The practical benefit for unit task performance is theoretically understandable because the intervention influenced the culminating task directly by clarifying rubric-based expectations and providing feedback. That is, the last activity was not a stand-alone evaluation but a component of the instructional design itself. The high inter-rater reliability in Table 3 supports this interpretation. More recent studies also indicate that rubric-based scoring may be useful where performance assessment criteria are clear and scoring reliability is high (Williamson, 2025), although more recent meta-analytic studies indicate that overall rubric-based scoring can be reliably moderate-high in terms of reliability and validity.

Table 6 of correlation gives additional insight into the pattern of findings. Learning engagement was more closely related to unit task performance than Chinese language competency was, which suggests that engagement may be particularly relevant for longer applied tasks that require sustained effort and involvement. This is in line with the engagement theory and with the more recent evidence that engagement plays a key role in the provision of learning outcomes (Fredricks et al., 2004; Xu et al., 2023).

Methodologically, the study also demonstrates a justifiable method of analysis of quasi experimental educational research. ANCOVA is often advised in case baseline measures are known since ANCOVA only enhances the accuracy of group comparisons, but it does not remove the overall constraints of nonrandomized designs (Vickers & Altman, 2001). When working with real classroom data in the future, it is also necessary to consider the possibility of class-level clustering, as well as to more systematically record treatment fidelity. In general, the current research is a consistent outline of assessing competency-based large-unit teaching using internationally accepted quantitative standards, but the findings are still tentative until tested on the real Sichuan data.

6.1. Implications

Theoretically, this study is useful to the literature in terms of demonstrating how Constructive Alignment, Assessment for Learning and Self-Determination Theory can be unified to describe the implications of competency-based large-unit teaching. Constructive Alignment assists in understanding why students can perform better in Chinese language competency and unit task performance when the desired outcomes, learning activities and unit assessment tasks are coherently connected at the unit level as opposed to being divided in disconnected lessons. Assessment for Learning goes further to explain the process by which such an instructional model can be effective, i.e. by providing continuous feedback, practice guided by rubrics, opportunities to revise, and visible success criteria which are built into the learning process. Simultaneously, Self-Determination Theory presents a motivational context explaining the more intense engagement in the experimental condition based on the idea that more specific goals, purposeful tasks, encouraging feedback, and learning in collaboration may be more beneficial in meeting the needs of autonomy, competence, and relatedness. Collectively, the research expands these theories by demonstrating that instructional coherence, formative assessment, and motivation-supportive design are not issues independent of each other, but rather, strengthening aspects of an integrated teaching-learning-assessment model.

Practically, the results indicate that teaching Chinese as the first language should not continue its traditional lesson-by-lesson delivery but rather be designed in concepts of more integrated units. The teachers are expected to plan reading, language use, expression, and performance activities around common competency objectives and to incorporate formative assessment in the unit in a variety of ways, including the use of rubrics, draft feedback, peer review, self-assessment, and revision checkpoints. The practices can not only enhance the academic competence of students but also enhance the engagement by making the learning objectives more explicit and visible. The findings at the school level suggest that effective use of large-unit instruction should be supported over time through unit planning, task design, rubric construction, and feedback practice as opposed to simply telling teachers to alter lesson format. Beyond that, the paper hypothesizes that reforming the curriculum is better received when the three issues of coherence, assessment, and student motivation are considered as a whole as a single instructional strategy.

7. Conclusion

This study aimed at investigating the possibility of competency-based large-unit instruction to yield superior results as compared to traditional lesson-by-lesson teaching in Grade 5 Chinese language classrooms. According to the results, the results were invariably in favor of the experimental group in all the three outcome variables. Students who were exposed to the integrated teaching-learning-assessment model showed improved post-test Chinese language competency, learning engagement, and unit task performance compared with students in the control group, but the adjusted results suggest moderate rather than large effects. This trend, as noted in the results section, could indicate that large unit instruction can be particularly successful when instructional objectives, learning tasks and assessment activities are packaged as an integrated whole and not as individual classroom units.

Conceptually, the findings can be related to the view that teaching and learning can be enhanced by focusing on curriculum alignment, formative assessment, and student motivation as a joint construct. Constructive Alignment assists clarify the reason as to why the coherent connections amid the planned outcomes, teaching activities and assessment undertakings can reinforced Chinese language competency and unit task achievement (Biggs, 1996; Wiggins & McTighe, 2005). Assessment for Learning also indicates that continuous feedback, revision and publicly visible success standards can be used to facilitate improvement throughout the learning process, not just at its completion (Black & Wiliam, 1998; Hattie & Timperley, 2007). Meanwhile, the moderate engagement effect can also be explained plausibly by Self-Determination Theory since meaningful tasks, constructive feedback, and related learning experience can be more effective in meeting the needs of students in autonomy, competence, and relatedness. The collective of the studies, in turn, confirms the opinion that competency-based large-unit teaching is not merely a question of a curriculum structure, but a more general pedagogical strategy that can also have a combined impact on cognitive, motivational, and performance-based results.

7.1. Limitations

There are a number of limitations to this study. First, the quasi-experimental design constrains causation since nonrandomized group assignment can cause selection bias, despite using baseline adjustment procedures such as ANCOVA (Vickers & Altman, 2001). Second, when the unit of assignment is intact classes, class-level clustering might impact on statistical estimates and can be considered in future research. Third, the study is not concerned with long-term outcomes, but only with short-term results. Lastly, the treatment fidelity could be different in different classrooms, and more systematic implementation monitoring should be incorporated in the future, to make the conclusions stronger.

References

- Andrade, H. L. (2013). Classroom Assessment in the Context of Learning Theory and Research. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 17–34). SAGE Publications. doi: <https://doi.org/10.4135/9781452218649.n2>
- Appleton, J. J., Christenson, S. L., Kim, D. & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology, 44*(5), pp. 427–445. doi: <https://doi.org/10.1016/j.jsp.2006.04.002>
- Arshad, M. Z. & Arshad, D. (2018). Intellectual Capital and SMEs Performance in Pakistan: The Role of Environmental Turbulence. *International Journal of Entrepreneurship, 22*, pp. 1–14. Retrieved from https://www.abacademies.org/articles/Intellectual-capital-smes-performance-in-pakistan_-the-role-of-environmental-turbulence-1939-4675-22-SI-189.pdf
- Berry, R. (2008). *Assessment for Learning* (Vol. 1). Hong Kong University Press. doi: <https://doi.org/10.1515/9789888052059>

- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), pp. 347–364. doi: <https://doi.org/10.1007/BF00138871>
- Biggs, J. & Tang, C. (2011). Train-the-Trainers: Implementing Outcomes-based Teaching and Learning in Malaysian Higher Education. *Malaysian Journal of Learning and Instruction*, 8, pp. 1–19. Retrieved from <https://e-journal.uum.edu.my/index.php/mjli/article/view/7624>
- Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), pp. 7–74. doi: <https://doi.org/10.1080/0969595980050102>
- Brookhart, S. M. (2013). *How to Create and Use Rubrics for Formative Assessment and Grading*. ASCD. Retrieved from <https://www.ascd.org/books/how-to-create-and-use-rubrics-for-formative-assessment-and-grading>
- Chan, C. K. Y. & Lee, K. K. W. (2021). Constructive alignment between holistic competency development and assessment in Hong Kong engineering education. *Journal of Engineering Education*, 110(2), pp. 437–457. doi: <https://doi.org/10.1002/jee.20392>
- Creswell, J. W. & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications. Retrieved from <https://edge.sagepub.com/creswellrd5e>
- Dann, R. (2014). Assessment as learning: blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, 21(2), pp. 149–166. doi: <https://doi.org/10.1080/0969594X.2014.898128>
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer Science & Business Media. doi: <https://doi.org/10.1007/978-1-4899-2271-7>
- Deci, E. L. & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology / Psychologie canadienne*, 49(3), pp. 182–185. doi: <https://doi.org/10.1037/a0012801>
- Deci, E. L. & Ryan, R. M. (2012). Self-determination theory. In *Handbook of Theories of Social Psychology: Volume 1* (pp. 416–437). SAGE Publications Inc. doi: <https://doi.org/10.4135/9781446249215.n21>
- Doo, M. Y. & Zhu, M. (2024). A meta-analysis of effects of self-directed learning in online learning environments. *Journal of Computer Assisted Learning*, 40(1), pp. 1–20. doi: <https://doi.org/10.1111/jcal.12865>
- Estes, T. S. (2016). *From the use of performance tasks to the user of performance tasks: Authentic learning and assessment experiences in middle schools* (Doctoral Dissertation, Aurora University). Retrieved from <https://www.proquest.com/openview/10d40c12a6497328b7a4cc5abba38229>
- Fang, H. & Xueyun, H. (2025). Core-Competence Educational Policies Versus the Practical Realities of Rural Area-Based Language Teachers: Pathways to Better Policy Implementation. *Sage Open*, 15(3), pp. 21582440251360144. doi: <https://doi.org/10.1177/21582440251360144>

- Fredricks, J. A., Blumenfeld, P. C. & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, 74(1), pp. 59–109. doi: <https://doi.org/10.3102/00346543074001059>
- Gao, S. & Pan, M. (2024). Implementing dynamic assessment in writing in Chinese as a second language: Towards an adjusted cumulative group dynamic assessment. *System*, 123, pp. 103327. doi: <https://doi.org/10.1016/j.system.2024.103327>
- Graves, K. & Garton, S. (2017). An analysis of three curriculum approaches to teaching English in public-sector schools. *Language Teaching*, 50(4), pp. 441–482. doi: <https://doi.org/10.1017/S0261444817000155>
- Guay, F. (2022). Applying Self-Determination Theory to Education: Regulations Types, Psychological Needs, and Autonomy Supporting Behaviors. *Canadian Journal of School Psychology*, 37(1), pp. 75–92. doi: <https://doi.org/10.1177/08295735211055355>
- Hao, J. (2024). Research on the Impact of Competency-Based Curriculum Design on Student Learning Outcomes. *Journal of Modern Educational Theory and Practice*, 1(2). doi: <https://doi.org/10.70767/jmetp.v1i2.272>
- Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), pp. 81–112. doi: <https://doi.org/10.3102/003465430298487>
- Heritage, M. (2007). Formative Assessment: What Do Teachers Need to Know and Do? *Phi Delta Kappan*, 89(2), pp. 140–145. doi: <https://doi.org/10.1177/003172170708900210>
- Hu, Y., Kantelinen, R. & Kautonen, M. (2024). Exploring the use of self-assessment in Chinese primary school EFL teaching: teachers' perspectives and practices in Shanghai, Shenzhen and Jiangsu. *Cogent Education*, 11(1), pp. 2425903. doi: <https://doi.org/10.1080/2331186X.2024.2425903>
- Jiang, M., Zhang, J. & Arshad, M. Z. (2025). Cultivating a creative ecosystem: the role of inclusive leadership in educational success. *Acta Psychologica*, 257, pp. 105076. doi: <https://doi.org/10.1016/j.actpsy.2025.105076>
- Joseph, S. & Juwah, C. (2012). Using constructive alignment theory to develop nursing skills curricula. *Nurse education in practice*, 12(1), pp. 52–59. doi: <https://doi.org/10.1016/j.nepr.2011.05.007>
- Kingston, N. & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*, 30(4), pp. 28–37. doi: <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Koo, T. K. & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), pp. 155–163. doi: <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lei, H., Cui, Y. & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal*, 46(3), pp. 517–528. doi: <https://doi.org/10.2224/sbp.7054>
- Lestari, P. (2025). Integrating Project-Based Learning within Outcome-Based Education Framework: A Literature Review on Polytechnic English Language Education. *Culturalistics: Journal of Cultural, Literary, and Linguistic Studies*, 9(1), pp. 32–36. doi: <https://doi.org/10.14710/ca.v9i1.27339>

- Li, J. (2022). Adaptive Learning Model of English Vocabulary Based on Blockchain and Deep Learning. *Mobile Information Systems*, 2022(1), pp. 4554190. doi: <https://doi.org/10.1155/2022/4554190>
- Li, J. & Gu, P. Y. (2026). Teacher continuing professional development in formative assessment: A pathway to enhanced student achievement. *Teaching and Teacher Education*, 174, pp. 105426. doi: <https://doi.org/10.1016/j.tate.2026.105426>
- Li, J., Xue, E., Li, C. & He, Y. (2023). Investigating Latent Interactions between Students' Affective Cognition and Learning Performance: Meta-Analysis of Affective and Cognitive Factors. *Behavioral Sciences*, 13(7), pp. 555. doi: <https://doi.org/10.3390/bs13070555>
- Liu, Y. (2024). The effect of classroom-based assessment (CBA) on Chinese EFL learners' speaking performance, engagement, and willingness to communicate: Undergraduate students' perceptions. *Heliyon*, 10(12), pp. e32629. doi: <https://doi.org/10.1016/j.heliyon.2024.e32629>
- Lou, S. S. (2025). *Analyzing the implementation of integrating teaching-learning-assessment (ITLA) in video-recorded English classroom practices in China* (Master's Thesis, University of Wisconsin–Platteville). Retrieved from <https://minds.wisconsin.edu/handle/1793/95280>
- Ministry of Education of the People's Republic of China. (2022a, April 21). *MOE releases compulsory education curriculum program and curriculum standards for Chinese and other subjects (2022 edition)*. Ministry of Education of the People's Republic of China. Retrieved from http://www.moe.gov.cn/srcsite/A26/s8001/202204/t20220420_619921.html
- Ministry of Education of the People's Republic of China. (2022b, April 21). *Notice on the issuance of the compulsory education curriculum program and curriculum standards (2022 edition)*. Retrieved from http://en.moe.gov.cn/news/press_releases/202205/t20220507_625532.html
- O'Donnell, C. L. (2008). Defining, Conceptualizing, and Measuring Fidelity of Implementation and Its Relationship to Outcomes in K–12 Curriculum Intervention Research. *Review of Educational Research*, 78(1), pp. 33–84. doi: <https://doi.org/10.3102/0034654307313793>
- OECD. (2019). *OECD Future of Education and Skills 2030. OECD Learning Compass 2030. A Series of Concept Notes*. Retrieved from <https://www.oecd.org/education/2030-project>
- Plengkham, B. & Wasanasomsithi, P. (2023). Effects of Integrated Performance Assessment Modules on English Speaking Ability of Thai EFL Undergraduate Students. *LEARN Journal: Language Education and Acquisition Research Network*, 16(1), pp. 448–472. Retrieved from <https://so04.tci-thaijo.org/index.php/LEARN/article/view/263451>
- Qi, W., Zhang, Y. & Arshad, M. Z. (2025). Sustainable education in action: Principal leadership, teacher wellbeing, and altruism in the sports context. *Acta Psychologica*, 254, pp. 104846. doi: <https://doi.org/10.1016/j.actpsy.2025.104846>

- Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y. & Pianta, R. C. (2016). How teacher emotional support motivates students: The mediating roles of perceived peer relatedness, autonomy support, and competence. *Learning and Instruction*, 42, pp. 95–103. doi: <https://doi.org/10.1016/j.learninstruc.2016.01.004>
- Ryan, R. M. & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, pp. 101860. doi: <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). Quasi-Experiments: Interrupted Time-Series Designs. In W. R. Shadish, T. D. Cook, & D. T. Campbell (Eds.), *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (pp. 171–205). New York: Houghton Mifflin. Retrieved from https://www.academicpeds.org/wp-content/uploads/gravity_forms/66-cffe81f82d0c22a20931a4634aea2e41/2023/03/Shadish-Cook-Cambell_2002_-ITS_Designs_OCR.pdf
- Shafii, R. A. & Berger, J.-L. (2025). Teacher assessment literacy, formative assessment practices, and their perceived efficacy in Tanzania: A scoping review. *Studies in Educational Evaluation*, 86, pp. 101496. doi: <https://doi.org/10.1016/j.stueduc.2025.101496>
- Sortwell, A., Trimble, K., Ferraz, R., Geelan, D. R., Hine, G., Ramirez-Campillo, R., et al. (2024). A Systematic Review of Meta-Analyses on the Impact of Formative Assessment on K-12 Students' Learning: Toward Sustainable Quality Education. *Sustainability*, 16(17), pp. 7826. doi: <https://doi.org/10.3390/su16177826>
- Tadesse, T., Ware, H., Asmare, A. & Gillies, R. M. (2024). Enhancing Student Engagement and Outcomes: The Effects of Cooperative Learning in an Ethiopian University's Classrooms. *Education Sciences*, 14(9), pp. 975. doi: <https://doi.org/10.3390/educsci14090975>
- Taras, M. (2007). Assessment for learning: understanding theory to improve practice. *Journal of Further and Higher Education*, 31(4), pp. 363–371. doi: <https://doi.org/10.1080/03098770701625746>
- Tashtoush, M. A., Shirawia, N. & Rasheed, N. M. (2025). Scoring Rubrics Method in Performance Assessment and its effect of Mathematical Achievement. *Athens Journal of Education*, 12(1), pp. 39–60. doi: <https://doi.org/10.30958/aje.12-1-3>
- Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, pp. 53–55. doi: <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Thanh, C. V. & Thao, T. Q. (2026). Revisiting Teacher's Roles under Competency-Based English Teaching Curriculum: A Systematic Literature Review. *VNU Journal of Science: Education Research*, 42(1), pp. 24–40. doi: <https://doi.org/10.25073/2588-1159/vnuer.5364>
- Thiel, O. & Hanssen, S. M. (2025). Engagement, study habits, and students' perceived competence in student-active early childhood teacher education. *Scandinavian Journal of Educational Research*, pp. 1–19. doi: <https://doi.org/10.1080/00313831.2025.2554729>
- Vickers, A. J. & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *BMJ*, 323(7321), pp. 1123–1124. doi: <https://doi.org/10.1136/bmj.323.7321.1123>

- Wang, J. B. (2025). *Exploring the implementation of “teaching-learning-assessment” integration in elementary school English teaching under the model of large-unit teaching* (Master’s Thesis, University of Wisconsin - Platteville). Retrieved from <https://minds.wisconsin.edu/handle/1793/95475>
- Wang, M. & Wei, R. (2025). Integrating Teaching-Learning-Assessment in English Reading Teaching of Chinese Junior High School Based on Backward Design. *International Journal of New Developments in Education*, 7(3), pp. 51–60. doi: <https://doi.org/10.25236/IJNDE.2025.070309>
- Wiggins, G. & McTighe, J. (2005). *Understanding by Design* (2nd ed.). ASCD. Retrieved from <https://www.ascd.org/books/understanding-by-design-expanded-2nd-edition?variant=103055>
- Williamson, B. (2025). Resiliency Resources for Lonely Readers: Teaching 18th-Century Environmental Lit through a Bibliotherapeutic Lens. *South Atlantic Review*, 90(1-2), pp. 115–135. Retrieved from https://southatlanticmla.org/wp-content/uploads/2025/09/SAR_90.1-2.pdf
- Wong, Z. Y., Liem, G. A. D., Chan, M. & Datu, J. A. D. (2024). Student engagement and its association with academic achievement and subjective well-being: A systematic review and meta-analysis. *Journal of Educational Psychology*, 116(1), pp. 48–75. doi: <https://doi.org/10.1037/edu0000833>
- Xu, X., Shi, Z., Bos, N. A. & Wu, H. (2023). Student engagement and learning outcomes: an empirical study applying a four-dimensional framework. *Medical Education Online*, 28(1), pp. 2268347. doi: <https://doi.org/10.1080/10872981.2023.2268347>
- Xuan, Q., Cheung, A. & Sun, D. (2022). The effectiveness of formative assessment for enhancing reading achievement in K-12 classrooms: A meta-analysis. *Frontiers in Psychology*, 13, pp. 990196. doi: <https://doi.org/10.3389/fpsyg.2022.990196>
- Yang, N. & Ghislandi, P. (2024). Quality teaching and learning in a fully online large university class: a mixed methods study on students’ behavioral, emotional, and cognitive engagement. *Higher Education*, 88(4), pp. 1353–1379. doi: <https://doi.org/10.1007/s10734-023-01173-y>
- Yang, S., Liu, L. & Hunt, N. (2022). Exploring the influence of perceived classroom environment on learner autonomy in a Chinese EFL learning context. *Frontiers in Psychology*, 13, pp. 1063473. doi: <https://doi.org/10.3389/fpsyg.2022.1063473>
- Zhang, H. (2025). Effect of Interaction and Alignment of Continuation Task in Chinese as a Second Language Writing. *Journal of Language, Identity & Education*, 24(1), pp. 1–13. doi: <https://doi.org/10.1080/15348458.2022.2072312>
- Zhao, L., Zhao, B. & Li, C. (2023). Alignment analysis of teaching–learning-assessment within the classroom: how teachers implement project-based learning under the curriculum standards. *Disciplinary and Interdisciplinary Science Education Research*, 5(1), pp. 13. doi: <https://doi.org/10.1186/s43031-023-00078-1>