INNOVATION IN LANGUAGE LEARNING International Conference

Integrating Core Literacy and Language Carriers: A Comparative Study of Biology Assessments in Chinese Gaokao and OxfordAQA A-Level

Ming Yan[1], Jing Gu[2], Longfei Wo[2*]

[1]Canadian Trillium College Nanjing Campus, PRC
[2]International School Affiliated to Nanjing University, PRC
[2]International School Affiliated to Nanjing University, PRC

Abstract

Biology assessments require students to explain mechanisms, handle data and use subject-specific language, but the balance between these demands differs across examination systems. This paper compares Gaokao Biology and OxfordAQA International A-Level Biology using public papers and curriculum materials from 2023 to 2025. Content analysis of task input, expected response and reasoning demand shows that Gaokao relies on Chinese-language contexts that require close reading, precise terminology and integration across concepts. OxfordAQA makes task requirements more explicit through command words and mark schemes, while placing greater demand on extended English explanation and evidence-based evaluation. Both systems can create difficulties for Chinese students, although the source of difficulty differs. A paired check with 36 students found stronger evidence use in English-medium tasks but lower scores for conceptual accuracy and academic-language control. Causal reasoning did not differ significantly. This paper uses these findings to propose a CLIL-based model that links concept mapping, academic language support and command-word reasoning.

Keywords: *Gaokao; OxfordAQA A-Level; Biology Assessment; Content and Language Integrated Learning (CLIL); Scientific Expression; Core Literacy*

1. Introduction

Current biology curricula place less emphasis on factual recall and more on what students can do with scientific knowledge. China's Biology Curriculum Standards set out life concepts, scientific thinking, inquiry and social responsibility as core goals [1]. International A-Level specifications express related aims through assessment objectives and command-word conventions [2]. The two systems share a broad view of biological competence, but students must show that competence through different forms of language and reasoning.

Scientific explanation in biology depends on precise terms, processes, conditions and causal links. A student may know the meaning of selective expression or homeostasis and still struggle to construct a mechanism that satisfies a particular mark scheme. Conceptual and linguistic demands therefore overlap in assessment, and separate scoring of content and language can obscure that overlap [3].

Chinese students in international or dual-curriculum programmes meet these demands in different ways. In the national system, they read dense Chinese stems and infer implicit conditions before providing short, precise answers. In an English-medium programme, they interpret command words, write at greater length and evaluate evidence in academic English. The same content knowledge may therefore appear stronger in one setting than another because the language and response demands have changed [4]. The paper asks how the two assessment systems organise these demands and what kind of instruction can help students move between them.

Research on comparative assessment has given considerable attention to content coverage and curriculum alignment. Less attention has been paid to how language organises the reasoning that examinations reward. This comparison treats each paper as a system for presenting information, eliciting answers and defining an acceptable scientific explanation. It is intended to identify patterns that may inform teaching; it does not rank the two qualifications or treat them as equivalent.

2. Theoretical and Analytical Framework

CLIL provides the main theoretical starting point. It treats language as part of disciplinary learning rather than a separate skill: students build biological meaning while learning how scientific explanations are worded and structured [5], [6]. Dual coding adds a role for diagrams and other visual representations

[7], and cognitive load theory helps explain why unfamiliar terminology combined with long response demands can interfere with performance [4]. Assessment theory connects these concerns by treating performance as an interaction between task, language ability and context [3].

The comparison uses three dimensions. Input refers to the density, explicitness and format of the material a student must interpret. Output refers to the precision, length and organisation of the response. Reasoning refers to the type of explanation, evaluation or judgement required. Separating these dimensions makes it possible to see which difficulties arise from biology, which from language, and where the two overlap.

Publicly available materials formed the documentary corpus. It included Gaokao Biology papers and curriculum standards from 2023 to 2025, together with OxfordAQA International A-Level Biology papers, specifications, mark schemes, examiner reports and unit materials [1], [2]. Curriculum and assessment objectives were mapped first. Task stems were then coded for context density, terminology, implicit conditions, scaffolding and visual data. Response demands were compared through question verbs, expected answer length, mark-scheme structure and the level of causal or evaluative reasoning. Examiner reports and scoring guidance were read for recurring language-related barriers. This part of the analysis describes assessment design and does not estimate item difficulty.

The comparison was covered with a paired task study involving 36 students in Grades 11-12. Each student completed three or four equivalent biology tasks in Chinese and English, with language order randomised or balanced. Two raters scored conceptual accuracy, causal reasoning, evidence use and academic language control on a 0-4 scale.

3. Results

3.1 Gaokao Biology: Context Density and Conceptual Precision

Gaokao Biology tasks often place biological problems within research, agriculture, ecology, health or everyday contexts. The contexts can be information-rich, and not every relevant condition is signposted. Students must decide what matters, link it to a core concept and express the answer with standard terminology. Because the curriculum develops a smaller number of central ideas, cross-module connections are also expected [1].

These tasks place pressure on information selection. Students have to separate essential conditions from background detail and keep several relationships in view while writing. When an unfamiliar context is paired with a familiar concept, the task tests transfer as well as reading. The main difficulty is rarely basic comprehension; it lies in recovering a precise biological relationship and expressing it within the available marks.

Precision is also demanded in the answer. Terms such as increase, decrease, promote, inhibit and selectively expressed can change the scientific meaning of a sentence, and qualifiers often determine whether a mark is awarded. Students are therefore assessed on both conceptual understanding and semantic control. Since the paper is in Chinese, decoding is not the main hurdle. The harder task is reading a dense stem carefully and expressing the relationship in the form the mark scheme expects.

3.2 OxfordAQA A-Level Biology: Command Words and Structured Reasoning

OxfordAQA questions use command words such as describe, explain, suggest, evaluate and justify. Each word signals a different cognitive operation and sets expectations for the evidence a response should contain. The framework separates knowledge and understanding, application, and analysis or evaluation, with experimental reasoning given particular weight [2].

Mark schemes reward either point-by-point content or broader levels of response. An explanation needs a cause, an intermediate process and a consequence. An evaluation requires comparison of evidence, recognition of limitations and a proportionate judgement. These conventions make the expected reasoning easier to identify, but the answer still has to be sustained in academic English. Students need command of verbs, connectors, hedging and transitions between claim and evidence; technical vocabulary alone is not enough. Examiner reports repeatedly draw attention to precision, sequence and evidence-based explanation [2].

OxfordAQA therefore reveals more of the reasoning pathway than Gaokao. Command words set the task, and mark schemes show how content earns credit. This transparency helps students who can read the task and write at length, but it can also reveal a gap between understanding the biology and being able to sequence an argument in a second language.

3.3 Cross-System Synthesis

The two systems place their demands in different places. Gaokao asks students to extract and integrate information from a dense Chinese context, then express the result with exact terminology and tightly controlled meaning. OxfordAQA reduces some uncertainty about what the question requires, but asks for more English, more extended reasoning and more explicit evaluation. Table 1 sets out the main differences, and Fig. 1 shows how language, input, response and reasoning demands interact.

Table 1. Comparison of Language and Assessment Demands in Two Biology Examination Systems

Dimension	Gaokao Biology (China)	OxfordAQA A-Level Biology (UK)
Language medium	Academic Chinese as a first language	Academic English as an additional language
Input features	Dense scientific contexts, implicit conditions, and rapid information selection	Explicit prompts, scaffolds, and comparatively transparent data cues
Output features	Precise short answers, controlled terminology, and integrated causal explanation	Structured extended responses, command-word logic, and level-based scoring
Language and reasoning	Lower first-language load with high interpretive precision and conceptual depth	Higher English lexical load with stepwise reasoning and explicit evaluation

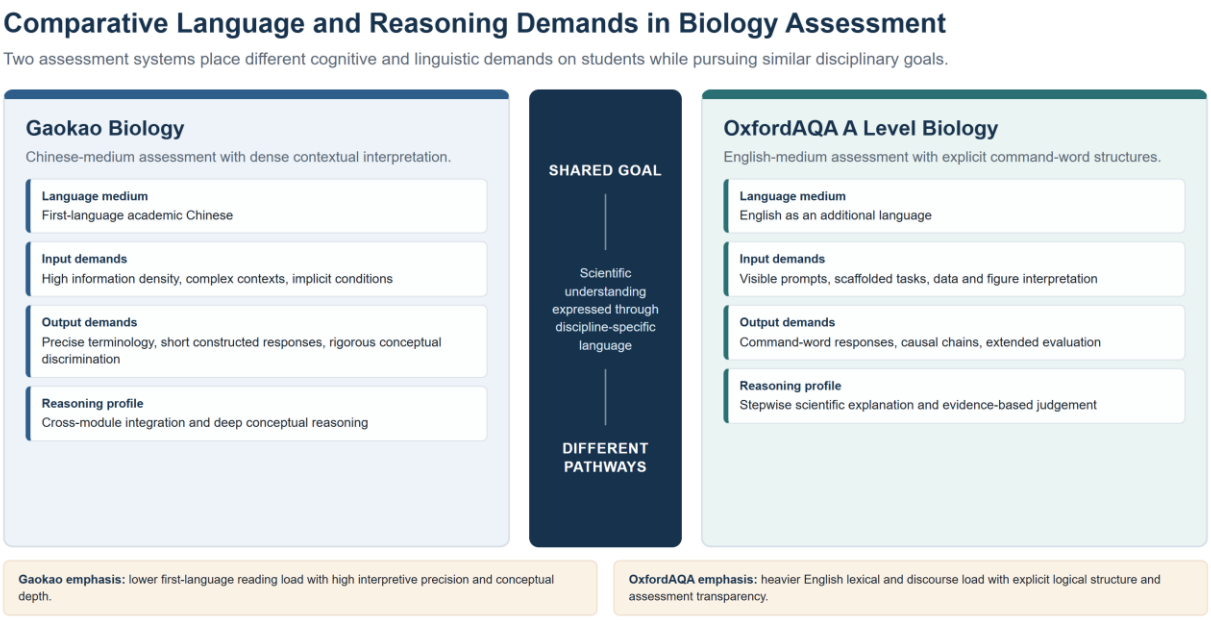


Fig. 1. Comparative Language and Reasoning Demands in Biology Assessment.

Neither pattern is inherently better. The Chinese examination gives students scope for conceptual integration and precise expression in their first language. The international examination makes explanation and evaluation more visible, although students must reach those demands through academic English. For students who move between the systems, the practical task is to gain control over both ways of reasoning.

3.4 Student Performance on Paired Tasks

The paired data did not show a uniform advantage for either language. English-medium tasks produced higher evidence-use scores (difference = 0.78, $p < .001$), but lower scores for conceptual accuracy (-0.21, $p < .001$) and academic language control (-0.90, $p < .001$). Causal reasoning was similar after Holm adjustment (-0.09, $p = .148$). Table 2 gives the full descriptive and inferential results.

Table 2. Paired Student Performance on Equivalent Chinese and English Biology Tasks

Dimension	Chinese M (SD)	English M (SD)	Difference (95% CI)	t	Holm p	dz
Conceptual accuracy	3.35 (0.19)	3.14 (0.25)	-0.21 [-0.32, -0.11]	-4.10	< .001	-0.68
Causal reasoning	3.01 (0.22)	2.92 (0.32)	-0.09 [-0.22, +0.04]	-1.48	.148	-0.25
Evidence use	2.42 (0.28)	3.20 (0.25)	+0.78 [+0.66, +0.90]	13.07	< .001	2.18
Academic language control	3.56 (0.15)	2.67 (0.25)	-0.90 [-0.98, -0.81]	-22.27	< .001	-3.71

Note. N = 36 paired students. Each language included three or four equivalent tasks scored from 0 to 4. Difference = English minus Chinese and was calculated from unrounded scores. Holm-adjusted p values are shown; paired-samples t tests, df = 35. dz = Cohen's d for paired samples.

4. Accessibility and Challenges for Chinese Learners

4.1 Gaokao Biology

Because Gaokao is taken in Chinese, initial decoding is less demanding. Familiar cultural, agricultural, ecological and scientific settings can help students bring prior knowledge into the task, and the paper rewards sustained conceptual reasoning. Problems arise when a stem brings together several facts, conditions and inferences. Constructed-response questions often give little direct guidance on organisation, so students have to work out both what to include and how to sequence it. A correct mechanism can still lose marks if a limiting word is omitted or the answer is pitched at the wrong level of specificity.

4.2 OxfordAQA A-Level Biology

Command words, data displays and staged questions make the demand of an OxfordAQA task easier to identify. Mark schemes also allow scientific logic to earn credit even when a minor language error does not change the meaning. The English-medium paper nevertheless creates lexical and discourse load. Students have to recognise long technical terms, interpret experimental language and move between data, explanation and judgement. Extended answers require a clear sequence, a distinction between evidence and interpretation, and conclusions that do not claim more than the data allow.

Evaluation questions add a further demand because students must weigh evidence for and against before reaching a judgement. Students who have mainly practised producing a final answer may need more help in building and qualifying an argument. The difference is not simply one of vocabulary: the assessment asks them to take part in a particular form of scientific discussion.

Classroom preparation therefore needs to respond to different reading and writing demands. Gaokao teaching needs close attention to context and semantic precision, but it should not turn reasoning into a search for wording cues. A-Level teaching needs explicit language support and repeated practice with causal and evaluative structures, without reducing biology to a set of answer templates. In both cases, students benefit when the reasoning behind an answer is made clear enough to use in a new question.

5. Discussion

The comparison shows that scientific literacy is assessed through a combination of content knowledge, language use and reasoning. Gaokao gives particular weight to semantic precision and conceptual integration. OxfordAQA gives particular weight to command-word interpretation, causal sequence and extended evaluation in English. Instruction that treats language as something added after the content has been taught is unlikely to prepare students adequately for either system.

Assessment design also indicates what students are expected to notice. A paper may reward quick reading of a dense context, exact terminology, a longer causal account or evaluation of competing evidence. Students adjust their study to these signals. Teaching can make those expectations explicit so that students understand the reasoning being assessed rather than only copying the surface format of past papers.

CLIL offers a useful way to bring content and language together [5], [6]. In biology, teachers can make disciplinary language visible and give students time to connect diagrams, terminology, mechanisms and evidence before asking for independent writing. Dual coding and multimodal translation provide ways to move between visual and verbal representations [7]. Command words should be taught as cognitive operations with different evidence requirements, not as labels to be memorised [3].

The paired tasks results are consistent with that approach. English performance was stronger for evidence use but weaker for conceptual precision and academic-language control, while causal reasoning was similar. The results support explicit teaching of disciplinary language and reasoning.

5.1 Integrated Instructional Model

The model begins with scientific meaning. Concept maps, mechanism diagrams and data displays give students a stable account of the biology before they write. Language support then follows the purpose of the task: sentence frames can help with causal explanation, comparison, limitation and judgement, but they should be removed as students gain control. Command words are grouped by the reasoning they require, and the same content is revisited as description, explanation or evaluation. The final element asks students to move between the two assessment traditions: Gaokao contexts develop inference and integration, while A-Level tasks develop experimental evaluation and evidence-based judgement. Fig. 2 presents the model.

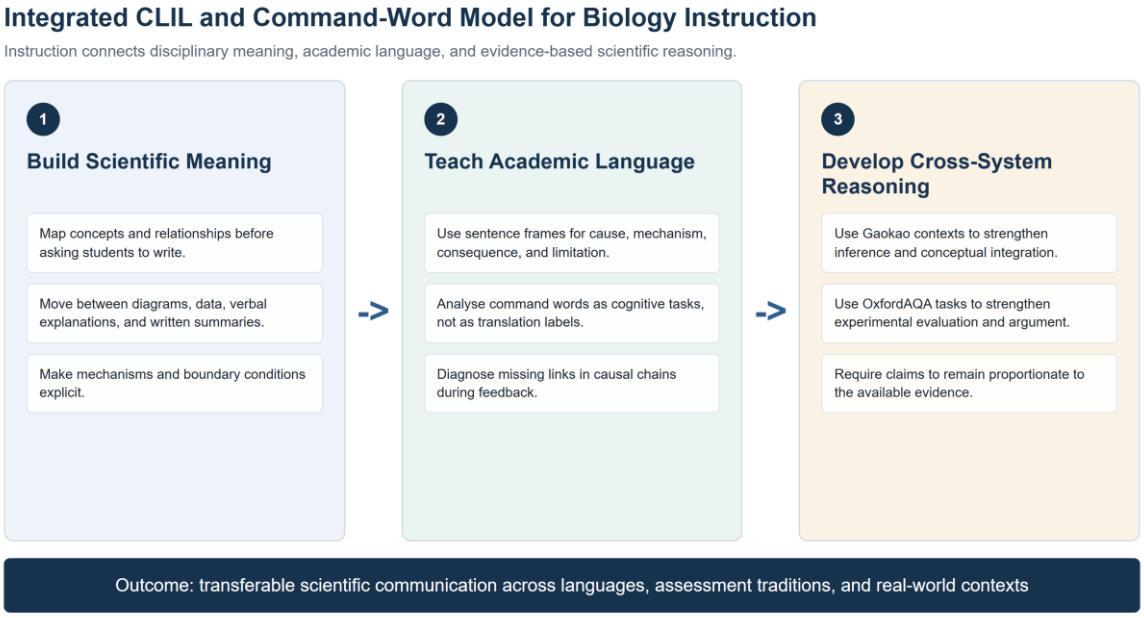


Fig. 2. Integrated CLIL and Command-Word Model for Biology Instruction.

6. Conclusion

Gaokao Biology and OxfordAQA A-Level Biology place different demands on students, even when the underlying biology overlaps. Gaokao rewards careful reading of dense Chinese contexts and precise conceptual expression. OxfordAQA rewards control of command words, extended causal explanation and explicit evaluative judgement in English. Students moving between the systems need practice in both, not a replacement of one approach by the other. The integrated CLIL and command-word model offers a way to make those demands explicit in teaching. The paired results point in the same direction: evidence use was stronger in English, while conceptual precision and academic language control were stronger in Chinese.

REFERENCES

[1] Ministry of Education of the People's Republic of China. (2020). Biology curriculum standards for senior secondary schools (2017 edition, 2020 revision). People's Education Press.
[2] OxfordAQA. (2023-2025). International A-Level Biology (9610) specification, question papers, mark schemes, and examiner reports. OxfordAQA.
[3] Bachmann, L., & Palmer, A. (2010). Language assessment in practice: Developing language assessments and justifying their use in the real world. Oxford University Press.
[4] Sweller, J., Ayres, P., & Kalyuga, S. (2011). Altering element interactivity and intrinsic cognitive load. In Cognitive load theory (pp. 203-218). New York, NY: Springer New York.
[5] Coyle, D. (2018). The place of CLIL in (bilingual) education. Theory Into Practice, 57(3), 166-176.
[6] Paran, A. (2013). CLIL: Content and Language Integrated Learning.
[7] Paivio, A. (1990). Mental representations: A dual coding approach. Oxford university press.