Reading Material Methodology for Heterogeneous ESL Classes

This document proposes a structured methodology for assigning reading material to ESL (English as a Second Language) students with vastly different proficiency levels, while maintaining a standardized data pipeline that feeds into the analytics and machine learning systems described in STUDENT_PERFORMANCE_ANALYSIS.md and STUDENT_PERFORMANCE_LLM.md.

Table of Contents

---

1. The Problem

1.1. Proficiency Heterogeneity

The ESL course receives students from three community college programs — Metrology, Biotechnology, and Cyber Security — with a single shared goal: improving English reading ability. The student population consistently splits into three proficiency bands:

Band	Approximate Share	Profile
Bottom third	~16 students	No prior connection to the English language. These students benefit from Duolingo for foundational exposure.

Band	Approximate Share	Profile
Middle third	~16 students	Studied English only in public high school. Can identify nouns, verbs, and adjectives in simple sentences. Limited productive vocabulary.
Top third	~16 students	Have studied in private English classes. Stronger foundations in grammar and vocabulary, some reading fluency.

1.2. The Current Approach

Reading material has been drawn from three sources of the student's choice:

- **Song lyrics:** Engaging but linguistically inconsistent (slang, ellipsis, non-standard grammar).
- **Short texts:** Varied quality and difficulty, no standardization.
- **Video transcripts:** Authentic material but difficulty varies wildly based on the video chosen.

Students create flashcards from these materials and review them using spaced repetition (Anki/StudyAmigo).

1.3. The Tension

Two goals pull in opposite directions:

- **Engagement** requires that material be interesting, relevant, and at the right difficulty level for each student — which means *different* material for different students.
- **Data standardization** requires that student performance be measured on *comparable* tasks — which means some *shared* material across all students.

The methodology below resolves this tension by structuring each exercise into three components: one shared, one tiered, and one free-choice.

2. Dual Placement: Duolingo + E01 Diagnostic

2.1. Rationale

Historically, students self-select their "road" (proficiency path) at the start of the course. This methodology formalizes and data-validates that self-assessment using two signals.

2.2. Signal 1: Duolingo Level (Pre-Semester)

For students who already use or are assigned Duolingo at the start of the course:

Duolingo Level	Suggested Initial Tier
Sections 1–2 (or no account)	Tier 1 (Foundation)
Sections 3–4	Tier 2 (Developing)
Section 5+	Tier 3 (Expanding)

For students without a Duolingo account, they are provisionally placed in Tier 1 until E01 data is available.

2.3. Signal 2: E01 as Diagnostic Exercise

E01 is redesigned as a **standardized diagnostic exercise** — the same content for all students, with no student-choice component. Its purpose is dual:

1. **Placement data:** The curated shared deck (Section 4) is distributed to all students. Their review performance (retention rate, average review time, ease distribution) after 1–2 weeks directly indicates proficiency level.
2. **Baseline metrics:** E01 establishes each student's baseline for all metrics tracked in STUDENT_PERFORMANCE_ANALYSIS.md, enabling meaningful measurement of improvement in subsequent exercises.

2.4. Combining Both Signals

After E01 completes, the instructor assigns final tier placement using a simple matrix:

	Low E01 Performance	Medium E01 Performance	High E01 Performance
No Duolingo / Low Duolingo	Tier 1 (confirmed)	Tier 1 or Tier 2 (instructor judgment)	Tier 2 (Duolingo underestimated)
Mid Duolingo	Tier 1 (Duolingo overestimated)	Tier 2 (confirmed)	Tier 2 or Tier 3 (instructor judgment)

	Low E01 Performance	Medium E01 Performance	High E01 Performance
High Duolingo	Tier 2 (Duolingo overestimated)	Tier 2 or Tier 3 (instructor judgment)	Tier 3 (confirmed)

E01 performance thresholds (to be calibrated after the first semester of data collection):

- **Low:** Retention rate < 50% on curated deck, or average review time > 20 seconds per card.
- **Medium:** Retention rate 50–75%, average review time 8–20 seconds.
- **High:** Retention rate > 75%, average review time < 8 seconds.

2.5. Tier Mobility

Tiers are not permanent. At the midpoint of the semester (after E04 or E05), students whose performance consistently exceeds or falls below their tier's expectations can be moved. The metrics from STUDENT_PERFORMANCE_ANALYSIS.md — particularly retention rate, improvement curve, and review density — inform this decision.

---

3. Three Proficiency Tiers

3.1. Tier 1 — Foundation

Target students: No prior English exposure or minimal passive exposure.

Reading characteristics:
– Individual sentences or 2–3 sentence micro-paragraphs.
– High-frequency vocabulary only (top 500 most common English words).
– Present tense exclusively in early exercises; past tense introduced gradually.
– Heavy use of Portuguese–English cognates (e.g., "important" / "importante", "technology" / "tecnologia") to build confidence.
– Visual support: images paired with text where possible.

Flashcard expectations:
– Simple word-translation pairs (English word → Portuguese meaning).
– Simple sentence → key word identification.
– Target: 5–10 new cards per exercise from tiered passages.

Parallel support: Continued Duolingo usage is encouraged to build foundational vocabulary and grammar outside the flashcard exercises.

3.2. Tier 2 — Developing

Target students: High school English background. Can parse simple sentences.

Reading characteristics:
– Short paragraphs (50–100 words).
– Compound and some complex sentences.
– Multiple tenses (present, past, future, present perfect).
– Vocabulary extends to top 1500 most common words plus domain-adjacent terms.
– Cognates still present but not relied upon.

Flashcard expectations:
– Sentence-level cards: English sentence → Portuguese translation or key concept.
– Vocabulary in context: English word used in a sentence → meaning.
– Grammar-focused cards: identifying verb tenses, sentence structures.
– Target: 10–15 new cards per exercise from tiered passages.

3.3. Tier 3 — Expanding

Target students: Private English class background. Can read simple texts with some fluency.

Reading characteristics:
– Full passages (150–300 words).
– Complex sentence structures (relative clauses, passive voice, conditionals).
– Academic and semi-technical vocabulary.
– Inferential comprehension required (what does the author mean, not just what do the words say).
– Minimal cognate scaffolding.

Flashcard expectations:
– Comprehension cards: question about the passage → answer.
– Vocabulary cards: word in context → definition and example usage.
– Grammar analysis: identifying and explaining complex structures.
– Target: 15–20 new cards per exercise from tiered passages.

---

4. The Curated Shared Deck

4.1. Replacing "1000 Basic English Words"

The "1000 Basic English Words" deck has served as the de facto shared baseline — nearly every student used it across all exercises in the 2024–2025 cohort. However, it was not designed for this population. The replacement deck should be purpose-built.

4.2. Design Principles

The curated shared deck should follow these principles:

1. **Cross-domain relevance:** Vocabulary drawn from themes common to Metrology, Biotechnology, and Cyber Security — science, technology, measurement, data, systems, safety — as well as general academic English.

2. **Internal tiering via tags:** Every card is tagged with a difficulty level (tier1, tier2, tier3). All students review tier1 cards. Tier 2 students also review tier2 cards. Tier 3 students review all three levels. This produces comparable data across tiers while respecting proficiency differences.

3. **Progressive release:** Cards are organized into weekly or bi-weekly batches aligned with the exercise schedule. New cards are introduced gradually rather than dumped all at once.

4. **Bilingual scaffolding:** Tier 1 cards include Portuguese hints or translations on the back. Tier 2 cards include English definitions with Portuguese fallback. Tier 3 cards are English-only with example sentences.

5. **Card type variety:**

   • Word → Definition (all tiers).
   • Sentence completion / cloze deletion (Tier 2 and 3).
   • Reading comprehension micro-questions (Tier 3).

4.3. Suggested Deck Size

Tier Level	Cards	Introduced Over
Tier 1 (Foundation)	300 cards	Exercises E01–E09 (~33 per exercise)
Tier 2 (Developing)	200 additional cards	Exercises E02–E09 (~25 per exercise)
Tier 3 (Expanding)	150 additional cards	Exercises E02–E09 (~19 per exercise)
Total deck	**650 cards**	

A Tier 3 student would be exposed to all 650 cards over the semester. A Tier 1 student would focus on the 300 foundation cards.

4.4. Content Sources for the Deck

- **General academic word list (AWL):** The 570 word families in Averil Coxhead's Academic Word List, filtered and adapted for Portuguese-speaking learners.
- **Cross-domain technical vocabulary:** Terms that appear across Metrology, Biotechnology, and Cyber Security (e.g., "accuracy", "protocol", "analysis", "system", "data", "control", "standard").
- **High-frequency English:** Oxford 3000 or New General Service List for foundation-tier words.
- **Cognate-aware selection:** Prioritize cognates early (Tier 1) to build confidence, then introduce false friends and non-cognates in higher tiers.

---

5. Exercise Structure: Three Components per Exercise

Each of the 9 exercises (E01–E09) is structured with three components. This design ensures that every exercise produces both standardized and personalized data.

5.1. Component A — Shared Deck Review (Standardized)

What: All students review their assigned portion of the curated shared deck.

Purpose: Produces directly comparable performance data across all students and tiers.

Data generated:
– Retention rate on shared cards (from revlog.ease).
– Review frequency and consistency (from revlog.id timestamps).
– Card maturity progression (from cards.ivl).

Weight in exercise grade: ~30%.

5.2. Component B — Tiered Reading Passage + Card Creation (Comparable)

What: The instructor provides a reading passage at three difficulty levels on the **same topic.** Students receive the version matching their tier and create flashcards from it.

Example topic: "How Vaccines Work"
– **Tier 1:** 3 simple sentences with labeled vocabulary. "A vaccine helps the body fight disease. The body learns to recognize the virus. This is called immunity."
– **Tier 2:** A 75-word paragraph with compound sentences. "When a person receives a vaccine, their immune system is exposed to a weakened form of the virus. The body produces antibodies, which are proteins that fight

the infection. If the person is exposed to the real virus later, their body remembers how to fight it."
– **Tier 3:** A 200–word passage with complex structures, passive voice, and technical vocabulary. Includes comprehension questions requiring inference.

Purpose: Allows cross–tier comparison on the same topic while respecting proficiency differences. Card creation quality can be evaluated against the known source text.

Data generated:
– Number and quality of cards created from the passage.
– Content analysis: do the cards capture key vocabulary and concepts?
– Comparison: how does card creation quality correlate with subsequent review performance?

Weight in exercise grade: ~40%.

5.3. Component C — Free–Choice Material (Personalized)

What: Students select their own reading material — song lyrics, video transcripts, articles, stories — and create flashcards from it.

Purpose: Maintains engagement and motivation. Produces rich, diverse data for the personalization ML models described in STUDENT_PERFORMANCE_ LLM.md.

Guardrails:
– Minimum card count requirement (e.g., 5 cards per exercise).
– Cards must be derived from a readable English source (not just word lists).
– Students submit the source URL or text alongside their cards.

Data generated:
– Student content preferences and self–selected difficulty level.
– Card creation patterns on unstructured material.
– Engagement metrics (do students review free–choice cards more or less than assigned cards?).

Weight in exercise grade: ~30%.

5.4. E01 Exception

E01 operates differently as a diagnostic exercise:

- **Component A only:** All students receive the first batch of the curated shared deck.
- **No Component B or C:** No tiered reading or free–choice material yet.
- **Purpose:** Pure diagnostic. Performance on Component A, combined with Duolingo level, determines tier placement for E02 onward.

5.5. Exercise Timeline Summary

Exercise	Component A (Shared Deck)	Component B (Tiered Passage)	Component C (Free Choice)
E01	Diagnostic batch (all students)	—	—
E02	Batch 2 + ongoing review	Passage 1 (first tiered reading)	First free-choice
E03	Batch 3 + ongoing review	Passage 2	Free-choice
E04	Batch 4 + ongoing review	Passage 3	Free-choice
E05	Batch 5 + ongoing review	Passage 4 (midpoint tier reassessment)	Free-choice
E06	Batch 6 + ongoing review	Passage 5	Free-choice
E07	Batch 7 + ongoing review	Passage 6	Free-choice
E08	Batch 8 + ongoing review	Passage 7	Free-choice
E09	Final review (all batches)	Passage 8 (summative)	Free-choice

6. Reading Passage Selection Criteria

6.1. Topic Selection: Cross-Domain Themes

Since all three programs (Metrology, Biotechnology, Cyber Security) share the same reading passages, topics should be drawn from the intersection of science, technology, and general knowledge. Suitable themes include:

- **Science and health:** How vaccines work, climate change basics, DNA and genetics, nutrition.
- **Technology and society:** Artificial intelligence, how the internet works, data privacy, renewable energy.
- **Measurement and standards:** Units of measurement, quality control, how sensors work.
- **Innovation stories:** Brief biographies of scientists/inventors, technology breakthroughs.

These topics have natural vocabulary overlap with all three technical programs without being exclusive to any one of them.

6.2. Grading Criteria by Tier

When creating three versions of the same topic, apply these measurable criteria:

Criterion	Tier 1	Tier 2	Tier 3
Text length	20–50 words (3–5 sentences)	75–150 words (1–2 paragraphs)	200–350 words (3–5 paragraphs)
Sentence length	5–8 words average	10–15 words average	15–25 words average
Vocabulary level	Top 500 frequency + cognates	Top 1500 frequency	Top 3000 frequency + academic
Grammatical structures	Present simple, imperative	Past simple, future, compound sentences	Passive voice, conditionals, relative clauses
Cognate density	High (>30% of content words)	Moderate (15–30%)	Low (<15%)
Comprehension type	Literal (what does the word mean?)	Literal + reorganization (summarize)	Inferential (why? what does the author imply?)

6.3. Readability Validation

Each passage should be validated using at least one readability metric:

- **Flesch–Kincaid Grade Level:** Tier 1 target grade 2–4, Tier 2 target grade 5–7, Tier 3 target grade 8–10.
- **Lexile measure:** If available, Tier 1 target 200–500L, Tier 2 target 500–800L, Tier 3 target 800–1100L.

These are guidelines, not rigid cutoffs. The Portuguese cognate density is a factor that standard English readability formulas do not account for — a passage full of cognates is effectively easier for Portuguese speakers than its Flesch–Kincaid score suggests.

6.4. Cognate Awareness for Portuguese Speakers

This population has a specific advantage: Portuguese and English share thousands of Latin-derived cognates. The methodology should exploit this strategically:

- **Tier 1:** Lean heavily on cognates to build confidence and demonstrate that English is more accessible than students assume. Words like "important", "problem", "information", "technology", "system" are immediately recognizable.

- **Tier 2:** Begin introducing false friends (words that look similar but have different meanings): "actually" ≠ "atualmente", "fabric" ≠ "fábrica", "pretend" ≠ "pretender", "push" ≠ "puxar". These make excellent flashcards.
- **Tier 3:** Use cognates less as scaffolding and more as a topic of linguistic analysis. Students at this level can discuss why cognates exist and identify them independently.

---

7. Connecting to the Data Pipeline

7.1. Three Components, Three Data Streams

The three-component exercise structure maps directly to the standardized data export format proposed in STUDENT_PERFORMANCE_LLM.md, Section 10:

Component	Deck Naming Convention	Data Pipeline Role
A (Shared Deck)	SHARED_CURATED_vX	Standardized cross-student comparison. Feeds difficulty estimation model and retention analysis.
B (Tiered Passage)	PASSAGE_E0X_TIER{1,2,3}	Tiered comparison. Feeds card quality scoring model. Content is known (instructor-provided), so card creation quality can be objectively assessed.
C (Free Choice)	Student-named (any name)	Personalization data. Feeds learning style clustering and engagement forecasting models.

By enforcing deck naming conventions for Components A and B, the data extraction scripts can automatically categorize and separate the three data streams without manual intervention.

7.2. Metrics Applicability by Component

The metrics from STUDENT_PERFORMANCE_ANALYSIS.md apply differently depending on the component:

Metric	Component A	Component B	Component C
Retention rate	Primary (standardized comparison)	Per-tier comparison	Individual tracking
Cards created (E0X)	N/A (pre-made deck)	Core metric (quality matters)	Volume metric
Review density	Consistency indicator	Engagement with passage	Intrinsic motivation proxy
Cramming ratio	SRS compliance	SRS compliance	Less critical (engagement-driven)
Ease factor distribution	Difficulty of shared content	Difficulty of tiered content	Student-selected difficulty
Card quality score	N/A (pre-made)	Assessable against source text	Harder to assess (no known source)

7.3. Exercise Schedule JSON Extension

The exercise schedule format from STUDENT_PERFORMANCE_LLM.md, Section 10.2 should be extended to include component information:

```
{
  "semester": "2025-2",
  "exercises": [
    {
      "id": "E01",
      "type": "diagnostic",
      "creation_start": "2025-08-04",
      "creation_end": "2025-08-04",
      "review_start": "2025-08-04",
      "review_end": "2025-08-18",
      "components": {
        "A": {
          "deck_name": "SHARED_CURATED_v1",
          "card_batch": "batch_01",
          "card_count": 33
        }
      },
      "notes": "Diagnostic exercise. No tiered passages or free-choice."
    },
    {
      "id": "E02",
      "type": "standard",
      "creation_start": "2025-08-18",
```

```
        "creation_end": "2025-09-01",
        "review_start": "2025-08-18",
        "review_end": "2025-09-08",
        "components": {
          "A": {
            "deck_name": "SHARED_CURATED_v1",
            "card_batch": "batch_02",
            "card_count": 33
          },
          "B": {
            "topic": "How Vaccines Work",
            "tier1_passage": "passages/E02_tier1.txt",
            "tier2_passage": "passages/E02_tier2.txt",
            "tier3_passage": "passages/E02_tier3.txt",
            "expected_cards_tier1": 5,
            "expected_cards_tier2": 10,
            "expected_cards_tier3": 15
          },
          "C": {
            "min_cards": 5,
            "source_submission_required": true
          }
        }
      }
    ]
}
```

7.4. Feeding ML Applications

Each component feeds specific ML applications from STUDENT_PERFORMANCE_
LLM.md:

ML Application	Primary Data Source
Early Struggling Student Detection	Component A (standardized retention data from E01)
Optimal Review Schedule Personalization	Components A + B (known-difficulty cards with review outcomes)
Difficulty Estimation for ESL Content	Component B (known passage text + measured student performance)
Learning Style Clustering	All three components (behavioral patterns across structured and unstructured tasks)
Flashcard Quality Scoring	Component B (student cards can be evaluated against the known source passage)
Automated Card Generation	Component B (pairs of passage text + high-quality student-created cards)

ML Application	Primary Data Source
Engagement Forecasting	Component C (free-choice engagement as intrinsic motivation signal)

---

8. Semester-over-Semester Improvement Loop

8.1. Semester 1 (Baseline)

- **Material selection**: Instructor manually creates or curates all passages and the shared deck.
- **Placement**: Duolingo + E01 diagnostic, with manual tier assignment.
- **Data collection**: Standardized export pipeline captures all three component streams.
- **Outcome**: First dataset of 40–50 students with structured, labeled interaction data.

8.2. Semester 2 (First Feedback)

- **Material selection**: Instructor still creates passages, but now informed by Semester 1 data:
  - Which Tier 1 passages were too hard? (Low retention rates across the tier.)
  - Which Tier 3 passages were too easy? (High ease factors, very short review times.)
  - Which topics generated the most engagement in Component C?
- **Placement**: Refined thresholds for E01 performance based on Semester 1 calibration.
- **Data collection**: Second cohort adds another 40–50 students to the dataset.
- **Outcome**: ~100 students of labeled data. Lightweight ML models (logistic regression for struggling student detection) become feasible.

8.3. Semester 3–4 (Model-Assisted)

- **Material selection**: Difficulty estimation model (trained on 100+ students of passage-performance pairs) can now predict whether a new passage is appropriate for each tier before it is assigned.
- **Placement**: E01 diagnostic can be partially automated — the model predicts tier placement from review patterns.
- **Card quality scoring**: Enough data to train a basic model that flags low-quality student-created cards.
- **Outcome**: ~200 students. Fine-tuning pre-trained language models on card content becomes viable.

8.4. Semester 5+ (Automated Generation)

- **Material selection:** The card generation model can propose flashcards from new reading passages, calibrated to each tier's proficiency level based on accumulated performance data.
- **Placement:** Near-automatic, validated by instructor review.
- **Personalized review schedules:** Per-student memory models adjust SRS intervals beyond SM-2 defaults.
- **Feedback loop complete:** Each semester's data improves the models that improve the next semester's material selection.

8.5. The Instructor's Evolving Role

Semester	Instructor Focus
1	Creating all materials, establishing the data pipeline, manual tier placement
2	Reviewing data insights, adjusting materials, refining tier thresholds
3–4	Reviewing model suggestions, curating (not creating) passages, focusing on intervention for at-risk students
5+	Supervising the system, focusing on pedagogical innovation, publishing research findings

The goal is not to replace the instructor but to shift their effort from content preparation toward student interaction and pedagogical design — the tasks that most benefit from human judgment and cannot be automated.