Diffit Quality Rubric

A scoring rubric for evaluating AI-generated instructional materials, derived from the Diffit Quality Constitution and the WestEd four-category framework for instructional materials quality (Bugler et al., 2017).

How to use this rubric

For each criterion, score the artifact:

- **Meets** — clearly satisfies the commitment on this artifact, with evidence
- **Partial** — partly satisfies; minor gaps a teacher could correct in seconds
- **Fails** — does not satisfy; a teacher would have to substantially rework

Every score must cite specific evidence — a quote, a question number, a page reference, a screenshot moment. Unsourced scores don't belong in this comparison.

N/A is honest. If a criterion doesn't apply to the prompt (e.g., source fidelity when no source was provided), mark N/A rather than forcing a score.

Diffit will not score perfectly. The Constitution is forward-looking; some commitments are aspiration. We score Diffit honestly and let the gap to competitors speak.

---

1. Accuracy and Craft

> The facts are right, the layout is clean, and a teacher will not be embarrassed in front of the class.

1.1 Factual correctness

Are facts, numbers, dates, names, units, and scientific claims accurate at a level a student at the named grade can verify? Does math compute?

Score	Anchor
Meets	Spot-check finds no factual errors in passage, vocab definitions, questions, or answer key
Partial	One minor, non-load-bearing inaccuracy (e.g., a date is the year before the actual year)
Fails	A factual claim a student or family could catch — wrong process, wrong number, wrong scientific mechanism

1.2 Source fidelity

When a source is provided (URL, PDF, text), does the material work *from* it without invention or substitution? Are quotations and data accurate to the source?

Score	Anchor
Meets	Every substantive claim traces to the source; no facts smuggled in from elsewhere
Partial	Mostly source-grounded; one or two details added or implied beyond the source
Fails	Invents details not in the source, fabricates quotes, substitutes a different source, or rewrites the problem
N/A	No source provided in the prompt

1.3 Mechanical correctness

Spelling, grammar, punctuation, capitalization. The boring stuff.

Score	Anchor
Meets	Clean enough to hand a student as a final draft
Partial	1–2 minor errors a teacher would not flag in a colleague's draft
Fails	Visible errors a teacher must correct before assigning

1.4 Layout discipline (page density, graphics, ink)

Does the page earn its length? Density, whitespace, font sizing scaled to grade. Graphics only when they do pedagogical work, not for decoration.

Score	Anchor
Meets	Density matches grade band; no padding; every graphic is informative
Partial	Usable but visually inefficient (e.g., 3 pages of what fits on 1; one decorative graphic)
Fails	Wall of text, padded with filler, or decorative graphics that distract

1.5 Answer-key consistency

Does the answer key actually answer the questions in *this* artifact, in the form the questions imply?

Score	Anchor
Meets	Every answer matches a question in this artifact; form (fraction/decimal, full sentence, etc.) matches
Partial	Most match; one or two minor drifts in form or numbering
Fails	Missing answers, wrong questions answered, or no key produced when one was expected
N/A	No answer key requested or implied

---

2. Standards Alignment and Depth of Knowledge

> The material hits the standard the teacher named, at the depth that standard requires — not a watered-down version of it.

2.1 Standards alignment

Is a specific standard cited (NGSS, CCSS, state-specific), and does the content actually meet it (not just "cover the topic")?

Score	Anchor
Meets	Explicit standard cited, and content addresses the verbs in the standard (e.g., NGSS 5-ESS2-1 asks for a model — the artifact has students build one)
Partial	Topic-correct, but no standard cited, or weakly aligned
Fails	No standard cited and content drifts off-topic, or claims a standard but does not meet its cognitive demand

2.2 Cognitive depth (DOK spread)

Across the questions in the packet, is there a real spread of cognitive demand, or all recall?

Score	Anchor
Meets	At least one question at DOK 3+ (analyze, evaluate, construct argument from evidence) alongside foundational recall
Partial	Mostly recall with one stretch question
Fails	All recall / one cognitive level only

2.3 Genuine variety

Do different question *formats* (MC, short answer, open-ended, matching) ask for meaningfully different *thinking*, or just the same recall in different visual layouts?

Score	Anchor
Meets	The thinking varies across questions, not just the format
Partial	Format varies; thinking varies somewhat
Fails	Same shallow recall asked three different ways

2.4 Question quality

Are multiple-choice distractors plausible, related to the content, and educative (revealing common misconceptions)? Does the position of the correct answer vary?

Score	Anchor
Meets	Distractors are content-relevant and educative; correct-answer positions vary across the set
Partial	One of the two: educative distractors but predictable positions, or varied positions but throwaway distractors
Fails	Throwaway distractors ("none of the above," nonsense options), or "always B" patterns

2.5 Honesty about what the standard requires

When a standard asks for extended writing, modeling, or argument from evidence, does the artifact actually demand that work — or is it satisfied with a fill-in-the-blank?

Score	Anchor
Meets	Where the standard demands depth, the artifact provides it (essay prompts, modeling tasks, evidence-based argument prompts)
Partial	Some depth, but the hardest cognitive moves are routed around
Fails	A standard requiring extended writing or modeling is satisfied with recall-level tasks
N/A	No standard cited

---

3. Ease of Use and Completeness

> A teacher can walk into the classroom and use this. No missing pieces, no internal inconsistencies, no "the teacher should now explain..." placeholders.

3.1 Multi-activity coherence

When the packet contains a passage + quiz + vocab + key, do they reference the same content — same details, same vocabulary, same emphasis?

Score	Anchor
Meets	The packet reads as a single resource; quiz asks about this passage; vocab drawn from this passage
Partial	Mostly coherent; minor drift (e.g., vocab not all drawn from passage)
Fails	Questions ask about content not in the passage; vocab unrelated; pieces feel independently generated

3.2 Complete and ready to teach

Can a teacher use the artifact as-is without hunting for missing pieces or filling in placeholders?

Score	Anchor
Meets	No placeholders, no "[insert passage here]," no "the teacher should now explain..."
Partial	Minor gaps a teacher fixes in under a minute
Fails	Stage directions, placeholders, or instructions to "find a source and fill it in"

3.3 Considered defaults

Does a single, simple prompt produce a usable output? Or does the teacher need to know the right follow-up prompts to get there?

Score	Anchor
Meets	One prompt, usable output. Early-career teacher could ship it
Partial	Usable with 1–2 obvious follow-ups
Fails	Requires significant prompt engineering / iteration to get to "usable"

3.4 Honest layout for student work

Does the space for student writing match what they're being asked to produce? (Essay → real lines; one-word answer → one line; show-your-work math → room for steps.)

Score	Anchor
Meets	Writing space appropriate to each task
Partial	One task miscalibrated (e.g., 3 lines for an essay)
Fails	Multiple tasks miscalibrated, or no writing space provided where it's needed

3.5 Classroom-workflow format fidelity

Does the output land cleanly in the formats teachers actually use — print-ready PDF, Google Docs, Google Slides, Forms?

Score	Anchor
Meets	Exports / prints cleanly to a teacher-ready file in at least two formats
Partial	Usable in one format; breaks in others
Fails	Copy-paste artifact; formatting collapses; no export path

---

4. Engagement and Meeting Student Needs

> The materials meet students where they are — reading level, language, prior knowledge, cultural context — and engage them rather than droning at them.

4.1 Right reading level

Sentence length, clause complexity, vocabulary, and conceptual density match the named grade.

Score	Anchor
Meets	Reads like the named grade — short clear sentences for younger, compound sentences and academic vocabulary (introduced with context) for older
Partial	Slightly above or below the named grade
Fails	Clearly miscalibrated — either dumbed-down (deleting content) or unreadable for the grade

4.2 Engaging voice

Does the passage hold a student's attention, or read as generic "AI-written" textbook drone?

Score	Anchor
Meets	Concrete examples, varied sentence structure, a voice a student would not skim past
Partial	Serviceable but flat
Fails	Generic, repetitive, or filler ("In this passage, we will learn about...")

4.3 Differentiation in the workflow

Can the same lesson be produced at multiple reading levels / languages / scaffolds from the same brief, without rebuilding?

Score	Anchor
Meets	Multiple differentiated versions delivered or one click away (e.g., 3 reading levels, ELL scaffold, translation)
Partial	Differentiation possible but requires separate, manually-crafted prompts
Fails	One version only; no path to differentiation in the workflow

4.4 Pedagogically-useful visuals

For topics where a diagram carries conceptual load (water cycle, photosynthesis, geometry, data interpretation), is the visual present, accurate, and grade-appropriate?

Score	Anchor
Meets	Diagram present, accurately labeled, scientifically correct, clarifies the concept
Partial	Present but generic, or partly inaccurate
Fails	No diagram for a topic that needs one; or a diagram that misrepresents the science / has hallucinated text artifacts

4.5 Asset-based framing and safety

Do examples, names, and contexts draw from broad cultural backgrounds rather than assuming a default student? Is content age-appropriate?

Score	Anchor
Meets	Examples draw from broad contexts; no default-student framing; nothing age-inappropriate
Partial	Neutral but bland; minor default-student assumptions
Fails	Excluding / stereotyping defaults, or age-inappropriate content

---

Scoring summary template

A side-by-side comparison should end with a heatmap like this:

Criterion	Diffit	Competitor
1.1 Factual correctness	✓	✓
1.2 Source fidelity	✓	⚠
…	…	…

Followed by:

- **Net assessment** — one paragraph
- **3–5 "look here" moments** — the specific evidence callouts an admin should look at

- **What we don't claim** — criteria where Diffit underperforms or where the difference is small. Honesty is the differentiator.

Citation

This rubric operationalizes the Diffit Quality Constitution, itself drawing on:

Bugler, D., Marple, S., Burr, E., Chen-Gaddini, M., & Finkelstein, N. (2017). *How Teachers Judge the Quality of Instructional Materials: Selecting Instructional Materials, Brief 1 — Quality.* WestEd. https://www.wested.org/resource/selecting-instructional-materials-brief-1-quality/

Diffit vs. Gemini for Education

A side-by-side comparison of two AI tools generating the same instructional packets across grade bands, scored against the Diffit Quality Rubric — which is derived from the Diffit Quality Constitution and WestEd's research on how teachers judge instructional-materials quality.

Audience: District and school leaders who already have Gemini for Education and want to understand what Diffit adds.

Approach: Three identical prompts, each run once in each tool, scored on observable artifact qualities. No marketing claims, no judgment calls about company stories — just what the tools produce when a teacher asks them for a reading packet, a math worksheet, or a primary-source lesson.

How this maps to the summary page: the four criteria shown there — *Classroom-ready & standards-aligned*, *Complete & ready to teach*, *Content integrity*, and *Differentiation that holds up* — are presentation themes drawn from the 20-criterion rubric. This document is the full per-criterion scoring behind them.

TL;DR

When a teacher asks each tool for the same packet, **Diffit produces a complete, classroom-ready resource in one prompt; Gemini produces a partial draft that needs follow-up prompting and still lacks structural elements teachers rely on.**

	Diffit	Gemini for Education
ELA packet (Harriet Tubman, NPS source, RI.5.3)	11 pages — full reading passage + 3 activities + answer key + 3 contextual images	3 pages — no passage, 3 questions, no answer key, no images. Required follow-up prompt to get even this
Math packet (fraction word problems, 5.NF.A.2)	15 pages — visual fraction models + word problems with estimation + error-analysis activity + full answer key	2–3 pages — 6 word problems + answer key. No visual fraction models. First response was unusable raw LaTeX
HS packet (MLK Letter from Birmingham Jail, RH.9-10.2)	17 pages — narrative context + matching tables + vocab analysis + crossword + MC quiz + short-answer + two answer keys	5 pages — narrative context + primary-source excerpts + 6 analysis questions + summary task with rubric. No answer key for the analysis questions
Considered defaults (one-prompt-to-usable)	✓ across all three prompts	✕ at elementary (required follow-up); ✓ at HS
Differentiation (Spanish + 2nd-grade relevel)	Complete packets in both variants (passage, activities, key, images preserved)	Workflow path exists but inherits structural gaps — variants still missing the reading passage and answer key
Pattern overall	Wins on completeness, activity variety, answer-key depth, source discipline across all grade bands	Closer to parity at HS than at elementary, but still fewer activities, weaker answer keys, and looser source attribution

The single most important finding: both tools were given the same brief; Diffit produced the full packet the brief asked for, and Gemini did not. This pattern shows up across both ELA and math, in different ways, and is driven by structural differences in how the two products are designed — not by prompt engineering.

We score Diffit honestly against its own (aspirational) Constitution and call out places where Diffit doesn't reach the bar — relevel oversimplifications, terminology inconsistencies in translation, missing visuals in the HS packet — so the wins it does post land with more weight.

What we tested

Three prompts spanning two grade bands, all CCSS-aligned, all core instructional needs. The two 5th-grade prompts (ELA + math) are detailed below. The HS prompt (MLK's Letter from Birmingham Jail at 9th–10th grade, aligned to RH.9-10.2) is detailed in the Pattern at secondary section.

Prompt 1 — ELA

> **Topic:** Harriet Tubman **Grade:** 5 **Standard:** CCSS.ELA-LITERACY.RI.5.3 — *"Explain the relationships or interactions between two or more individuals, events, ideas, or concepts in a historical, scientific, or technical text based on specific information in the text."* **Source:** NPS — Harriet Tubman Underground Railroad National Historical Park

Why this prompt: Universally taught in 5th-grade ELA and social studies. RI.5.3 is a workhorse standard with teeth — a packet that just summarizes Tubman's life does not meet it. The NPS source is unimpeachable authority and written at an adult reading level (real relieveling test).

In Diffit: Topic + grade + source URL + selected standard (CCSS.ELA-LITERACY.RI.5.3 via Standards picker). No follow-ups.

In Gemini: Single free-text message containing the equivalent brief.

Prompt 2 — Math

> **Topic:** Adding and subtracting fractions with unlike denominators (word problems) **Grade:** 5 **Standard:** CCSS.MATH.CONTENT.5.NF.A.2 — *"Solve word problems involving addition and subtraction of fractions referring to the same whole, including cases of unlike denominators, e.g., by using visual fraction models or equations to represent the problem."* **Source:** none (the standard is the brief)

Why this prompt: Fractions are the 5th-grade math focal point. 5.NF.A.2 explicitly demands **visual fraction models** and **word problems** — both areas where AI math output commonly falls short. Pure default-generation test.

The primary artifacts

Artifact	File	Pages
Diffit ELA — Harriet Tubman	diffit-ela-tubman.pdf	11
Gemini ELA — first response (default)	gemini-ela-tubman-first-response.pdf	2
Gemini ELA — after follow-up "make it a student handout"	gemini-ela-tubman.pdf	3
Diffit Math — fractions	diffit-math-fractions.pdf	15
Gemini Math — first response (default)	gemini-math-fractions-first-response.pdf	2
Gemini Math — after follow-up "fix the rendering"	gemini-math-fractions.pdf	3

Note: **In both subjects, Gemini's first response required a follow-up prompt to become usable.** That's a §3.3 (considered defaults) finding by itself.

ELA scoring

Diffit ELA vs Gemini ELA (final iterated version), against the rubric:

#	Criterion	Diffit	Gemini	Evidence
1.1	Factual correctness	✓ Meets	✓ Meets	All facts in both artifacts (dates, places, names, numbers) verified against the NPS source page
1.2	Source fidelity	✓ Meets	✓ Meets	All quoted phrases in both artifacts appear in the NPS source (mariners' mobility, marshlands, network of communication, 13 trips, 70+ rescued, etc.)
1.3	Mechanical correctness	✓ Meets	✓ Meets	Both clean of spelling/grammar errors
1.4	Layout discipline	✓ Meets	⚠ Partial	Diffit: paced for 5th grade with embedded images that earn their space. Gemini: dense bullet lists designed as teacher analysis, not student-paced
1.5	Answer-key consistency	✓ Meets	✕ Fails	Diffit includes 2-page answer key with sample paragraph responses and per-row table answers. Gemini includes no answer key at all
2.1	Standards alignment	✓ Meets	⚠ Partial	Diffit: activities directly target RI.5.3's verb (relationship analysis between individuals, events, ideas) and operate on the included passage. Gemini: the standard requires analysis "based on specific information in the text," but the packet asks students to analyze a passage that is not included — there is no text for the student to base their answers on
2.2	Cognitive depth	✓ Meets	⚠ Partial	Diffit: three activities spanning recall (cause/effect table), application (interactions table), and evaluation (extended writing prompt). Gemini: three open-ended "explain" questions — analysis-level but no spread
2.3	Genuine variety	✓ Meets	⚠ Partial	Diffit: three structurally different activities each demanding distinct cognitive work. Gemini: three questions all asking students to "explain"
2.4	Question quality	N/A	N/A	Primarily open-ended in both; not exercised
2.5	Honesty about what the standard requires	✓ Meets	✓ Meets	Both engage with relationship analysis; neither defaults to recall
3.1	Multi-activity coherence	✓ Meets	✕ Fails	Diffit: every activity references the included passage by specific details and vocabulary. Gemini: no passage exists for activities to cohere around — the questions reference a source the student doesn't have
3.2	Complete and ready to teach	✓ Meets	✕ Fails	Diffit: passage + 3 activities + answer key + Name/Date lines. Gemini: no reading passage and no answer key

#	Criterion	Diffit	Gemini	Evidence
3.3	Considered defaults	✓ Meets	✕ Fails	Diffit: one prompt → 11-page packet. Gemini: first response was a teacher-facing analysis; user had to explicitly prompt "turn it into a student handout" to get even the 3-page version
3.4	Honest layout for student work	✓ Meets	⚠ Partial	Diffit: writing space appropriate to each task (table cells, paragraph lines, reflection space). Gemini: limited writing space, designed more for analysis than student work
3.5	Classroom-workflow format fidelity	✓ Meets	✕ Fails (default) → ✓ Meets (iterated)	Diffit: clean PDF, prints and exports cleanly. Gemini default: includes raw LaTeX `$$\text{Death of Enslaver} \rightarrow \text{Financial Debt}$$` that did not render in Google Docs export. See gemini-ela-tubman-first-response.pdf page 2
4.1	Right reading level	✓ Meets	⚠ Partial	Diffit: passage reads cleanly at 5th-grade level — short sentences, age-appropriate vocabulary, structured headings. Gemini: no continuous passage to evaluate; analytical commentary uses sentence structures above 5th grade
4.2	Engaging voice	✓ Meets	⚠ Partial	Diffit: narrative passage with concrete details (the two-pound weight, the muskrat traps, "never lost a passenger"). Gemini: analytical, teacher-facing tone
4.3	Differentiation in the workflow	N/A	⚠ Partial	Tested in the Differentiation in practice section below (N/A here for Diffit).
4.4	Pedagogically-useful visuals	N/A	✕ Fails	A biography reading does not require a conceptual diagram, so this is N/A for Diffit (which nonetheless includes contextual images). Gemini produced no image either.
4.5	Asset-based framing	✓ Meets	✓ Meets	Both handle the subject matter respectfully and accurately

ELA scorecard: Diffit 17 Meets, 0 Partials, 0 Fails (out of 17 applicable; §2.4, §4.3, §4.4 N/A for Diffit). Gemini 4 Meets, 8 Partials, 5 Fails (out of 17 applicable, default version 1 additional Fail on §3.5). The §4.3 row is updated based on real differentiation artifacts; see the Differentiation in practice section below.

Math scoring

Diffit Math vs Gemini Math (final corrected version, with the default also noted), against the rubric:

#	Criterion	Diffit	Gemini (default)	Gemini (corrected)	Evidence
1.1	Factual correctness	✓ Meets	✓ Meets	✓ Meets	All computations verified: 3/4 + 1/3 = 13/12 ✓, 5/6 + 2/5 = 37/30 ✓, all Fraction Freddy corrections valid, Gemini's 6 word problems all compute correctly
1.2	Source fidelity	N/A	N/A	N/A	No source provided
1.3	Mechanical correctness	⚠ Partial	✕ Fails	Gemini is clean. Diffit's error-analysis sheet carries a stray rendering artifact — "½ + $14" where "½ + ¼" was meant — a minor, non-load-bearing slip, scored honestly.	
1.4	Layout discipline	✓ Meets	✕ Fails	⚠ Partial	Diffit: structured into three activity types with clear visual scaffolding. Gemini corrected: clean but minimal — no visuals, no scaffolding structure
1.5	Answer-key consistency	✓ Meets	⚠ Partial	✓ Meets	Diffit: full key including reasoning. Gemini default: key present but rendered in unreadable LaTeX. Corrected: clean answer key table
2.1	Standards alignment	✓ Meets	⚠ Partial	✓ Meets	Diffit: addresses 5.NF.A.2 fully — visual fraction models, word problems with unlike denominators, estimation, reasoning. Gemini default: alignment unverifiable due to broken LaTeX rendering. Corrected: addresses the topic via word problems
2.2	Cognitive depth	✓ Meets	✕ Fails	✕ Fails	Diffit: procedural (visual modeling) + application (word problems + estimation) + analysis (Fraction Freddy error-analysis). Gemini: six word problems, all single-cognitive-demand
2.3	Genuine variety	✓ Meets	✕ Fails	✕ Fails	Diffit: three structurally different activity types. Gemini: same task six times

#	Criterion	Diffit	Gemini (default)	Gemini (corrected)	Evidence
2.4	Question quality	N/A	N/A	MC quality is N/A for Diffit (its math is open-response).	
2.5	Honesty about what the standard requires	✓ Meets	✕ Fails	✕ Fails	5.NF.A.2 explicitly requires **visual fraction models or equations to represent the problem**. Diffit includes visual fraction models. Gemini includes neither — just text word problems
3.1	Multi-activity coherence	✓ Meets	N/A	N/A	Diffit: three activities all build on the same conceptual scaffold (finding common denominators, reasoning about reasonableness). Gemini: only one activity exists
3.2	Complete and ready to teach	✓ Meets	✕ Fails	⚠ Partial	Diffit: full instructional packet with all parts. Gemini default: unusable due to LaTeX. Corrected: usable but minimal — a teacher gets one practice activity, not a packet
3.3	Considered defaults	✓ Meets	✕ Fails	✕ Fails	Diffit: one prompt → 15-page packet. Gemini: required follow-up to fix broken rendering. The "default" experience produced an artifact a teacher could not give a student
3.4	Honest layout for student work	✓ Meets	⚠ Partial	⚠ Partial	Diffit: dedicated work space per problem matched to the steps required (estimation line + solve box). Gemini: a single "Work Space:" line per problem
3.5	Classroom-workflow format fidelity	✓ Meets	✕ Fails	✓ Meets	This is the clearest format-fidelity contrast in the comparison. Gemini default: **the entire worksheet is rendered as raw LaTeX source code**. See p1 of gemini-math-fractions-first-response.pdf. Gemini self-claims in its first sentence: "here is a clean, ready-to-print worksheet" — it literally is not. Corrected version (after follow-up) renders cleanly

#	Criterion	Diffit	Gemini (default)	Gemini (corrected)	Evidence
4.1	Right reading level	✓ Meets	✓ Meets	✓ Meets	Both produce word problems at 5th-grade reading level
4.2	Engaging voice	✓ Meets	✓ Meets	✓ Meets	Diffit: baking, hiking, gardening, pizza, track meet — diverse contexts. Gemini: reading, recipe, ribbon, pizza, puppy, water — also varied
4.3	Differentiation in the workflow	N/A	⚠ Partial	Tested in the Differentiation in practice section below (N/A here for Diffit).	
4.4	Pedagogically-useful visuals	✓ Meets	✕ Fails	✕ Fails	**5.NF.A.2 explicitly demands visual fraction models.** Diffit includes them on the first 4 pages of the packet (circles and bars). Gemini: zero visuals in either version. A teacher who hands a student the Gemini worksheet does not give them the visual scaffold the standard requires
4.5	Asset-based framing	✓ Meets	✓ Meets	✓ Meets	Diffit names: Maya, Leo, Sarah, Chloe, "Fraction Freddy." Gemini names: Lily, Marcus, Olivia, Liam, Noah, Sophia. Both reasonably diverse

Math scorecard: Diffit 16 Meets, 1 Partial, 0 Fails (out of 17 applicable; §1.2, §2.4, §4.3 N/A — the Partial is the stray-"$" rendering artifact on the error-analysis sheet, §1.3). Gemini corrected: 7 Meets, 5 Partials, 5 Fails. Gemini default: 1 Meets, 2 Partials, 8 Fails — an artifact a teacher could not give to a student without rebuilding it first.

Differentiation in practice

The ELA scoring above marked **§4.3 (differentiation in workflow)** as Meets for Diffit on faith — the rubric noted *"based on documented Diffit workflow capabilities; not exercised in this artifact run."* This section exercises it directly.

Starting from the same 5th-grade Tubman packet, both tools were asked to produce two differentiated versions: - **A Spanish translation** preserving formatting - **A 2nd-grade relevel** — two grade levels lower

than the original

In Diffit: each variant was one button click in the workflow (Translate → Spanish; Relevel → 2nd grade). In Gemini: each variant was one follow-up message in the same chat (*"Now translate this to Spanish, preserving formatting"* | *"Now make this for 2nd grade students"*). Both paths are workflow-cheap. The question is what each path produces.

The four differentiated artifacts

Variant	Diffit	Gemini
Spanish translation	diffit-ela-tubman-spanish.pdf — 10 pages	gemini-ela-tubman-spanish.pdf — 3 pages
2nd-grade relevel	diffit-ela-tubman-2nd-grade.pdf — 10 pages	gemini-ela-tubman-2nd-grade.pdf — 3 pages

Side-by-side: what survives the transform

Aspect	Diffit Spanish	Diffit 2nd grade	Gemini Spanish	Gemini 2nd grade
Pages	10	10	3	3
Reading passage	✓ (full, 4 pages)	✓ (full, 3 pages)	×	×
Three distinct activities	✓	✓	× (3 open Qs)	× (3 open Qs)
Answer key	✓ (2 pages)	✓ (2 pages)	×	×
Images preserved	✓ (same 3, identical files)	✓ (same 3, identical files)	×	×
Named historical figures retained	✓ Brodess, Rit, Ben, John Tubman, Andrew, Montgomery	✓ All preserved	n/a (no passage)	× Brodess removed; no Andrew, Montgomery
Specific dates and places	✓ 1822, 1849, 1862, 1863, Bucktown, Parson's Creek, Philadelphia, Combahee	✓ All preserved	n/a	× Most dropped; only "1849" and "13 times" survive
Combahee Raid / military leadership	✓	✓ (with one Partial finding, below)	n/a	× Entire section removed

The headline finding

Differentiation produces a version of whatever the tool produced originally. It does not fix structural gaps.

Diffit's original 5th-grade packet was a complete reading lesson — passage, three activities, answer key, three images. Its differentiated versions are also complete reading lessons in Spanish and at 2nd-grade level. Same passage in two languages and two reading levels, same activities, same key, same visuals.

Gemini's original was an analysis worksheet without a reading passage. Its differentiated versions are also analysis worksheets without reading passages. Translation didn't add a passage. Relieveling didn't add a passage. The gap inherits.

This is the §4.3 argument the original scoring made on faith. We now have artifacts. The workflow capability exists in both tools — that's not the differentiator. **The differentiator is what the workflow operates on.** When Diffit's underlying packet is complete, the differentiated variants are complete. When Gemini's underlying packet is missing structural elements, the differentiated variants are missing the same elements.

How the two tools approach relieveling

Going 5th → 2nd is a hard test. The Constitution names a specific commitment: *"We do not 'dumb down' by deleting content. We find the simpler way to express the harder idea. A relevel returns roughly the same amount of content, organized in roughly the same way."*

Diffit's relevel preserves content while simplifying expression. Compare these passages.

5th-grade original:

> *"From a very young age, her life was shaped by the hardships of slavery and the pain of being kept away from her loved ones."*

Diffit's 2nd-grade relevel:

> *"When she was little, she was a slave. This was very hard for her. She was also sad because she was away from her family."*

One complex sentence becomes three single-clause sentences. Same idea, simpler expression. Edward Brodess is still named in the 2nd-grade version. Rit and Ben are named. The two-pound weight incident is there. Parson's Creek is there. The Combahee River Raid is there. The 750 freed are in the answer key. Three images are preserved at identical file sizes.

Gemini's relevel removes content to achieve simplification. Edward Brodess becomes *"the man who owned Harriet."* Rit, Ben, Bucktown, the Little Blackwater River, Parson's Creek, Philadelphia — all gone. The entire fourth section of the original — Tubman's military leadership, the Combahee River Raid, Colonel James Montgomery, the 750 freed enslaved people — is absent from the 2nd-grade relevel. Gemini's 2nd-

grade version is shorter and less factually rich than its 5th-grade version, not because the expression got simpler, but because the content got cut.

This is the specific failure mode the Constitution's §4 names. Diffit takes one path; Gemini takes the other.

Honest findings on each side

These are real. We include them because honesty is what makes the wins above credible.

Against Diffit's 2nd-grade relevel: - Problem 4 in the Cause/Effect activity flattens *"plans and leads an armed mission along the Combahee River during the Civil War"* to *"helps soldiers in a war. They go on a boat trip."* The 750 freed are still in the answer key, but a 2nd grader reading just this prompt would have a substantially wrong mental model of what Tubman did. This is **§4-x meaning-preserving adaptation Partial** on a single activity item, not on the relevel as a whole. - "Confederate supply lines" becomes "enemy supply lines" — loses historical specificity. - The relevel uses "slave" where the original used "enslaved" (some districts have specific guidance on this terminology). - Title shift from *"The Life and Legacy of Harriet Tubman"* to *"Harriet Tubman: A Special Lady"* — "Special Lady" is twee.

Against Diffit's Spanish translation: - Mild terminology inconsistency on "marshlands": the reading passage uses *"los pantanos de Maryland"* but Activity 1 uses *"Las Marismas de Maryland"*. Both are valid, but a student looking for the exact phrase from the passage could be tripped up. - One untranslated phrase in the answer key: *"vendidas al Deep South"* — everywhere else the document uses *"Sur Profundo"*.

Against Gemini's Spanish translation: - Section 3 heading reads *"Decisión Cambia-Vidas"* — non-standard Spanish construction. The natural phrasing would be *"Decisión que Cambia Vidas"* or *"Decisión Trascendental."*

Against Gemini's 2nd-grade relevel: - The structural cuts named above (named figures dropped, military leadership section absent) are the §4-x finding. - Title softening to *"Harriet Tubman: Making Connections!"* with an exclamation point — twee in the same way Diffit's "Special Lady" is, but milder.

What this updates in the scoring

§4.3 differentiation in workflow was previously scored on faith for both tools. With artifacts in hand:

- **Diffit: Meets.** Workflow click produces differentiated versions that preserve the complete packet structure — passage, activities, answer key, visuals — in both Spanish and at 2nd grade. Honest single-activity finding noted on the relevel.
- **Gemini: Partial.** The workflow path exists and operates cheaply (one follow-up message per variant). But the differentiated outputs inherit the structural gaps of the original packet — no reading passage, no answer key, no images. A teacher needing differentiation in their workflow still has to manually add the missing structural elements themselves.

Pattern at secondary: Letter from Birmingham Jail

A reasonable objection to the comparisons above: *"Fine for elementary — but does the pattern hold at the secondary level, where the source is at the target reading level and the rubric exercises different muscles?"*

This section tests it. Same workflow as before, with a high-school prompt designed to exercise primary-source analysis at academic register — a different test than the elementary Tubman packet, which leaned on relieveling and historical narrative.

What we tested

> **Topic:** Martin Luther King Jr.'s "Letter from Birmingham Jail" **Grade:** 9–10 **Standard: CCSS.ELA-LITERACY.RH.9-10.2** — *"Determine the central ideas or information of a primary or secondary source; provide an accurate summary of how key events or ideas develop over the course of the text."* **Source:** Stanford King Institute encyclopedia entry on the letter

This standard has teeth: it demands not just identifying ideas but tracing *how ideas develop over the course of the text*. A packet that just summarizes the letter does not meet it.

Artifacts

Tool	File	Pages
Diffit	diffit-ela-mlk.pdf	17
Gemini for Education	gemini-ela-mlk.pdf	5

Diffit also offered two optional follow-up questions in the UI ("How should students engage with this text?" and "Which focus area do you want to emphasize for the summary?"). **Both were skipped** to keep this a true "default behavior" test. The 17-page output is what Diffit produces when a teacher selects topic, grade, standard, and source, then presses generate without configuring anything else.

Scoring

#	Criterion	Diffit	Gemini	Evidence
1.1	Factual correctness	✓ Meets	✓ Meets	All facts in both artifacts verified against canonical knowledge of the letter (April 12 1963 arrest, eight clergy, "unwise and untimely," composition on newspaper scraps, four steps, Why We Can't Wait, etc.)
1.2	Source fidelity	✓ Meets	⚠ Partial	Diffit attributes everything to the Stanford King Institute. Gemini's "Sources" section lists 10 URLs including govinfo.gov, Wikipedia, and the Beaverton OR city government — the prompt specified one source. Facts are real; the source-fidelity bar from the Constitution (*"We do not smuggle in facts from elsewhere, even when those facts are true"*) is not met
1.3	Mechanical correctness	✓ Meets	✓ Meets	Both clean
1.4	Layout discipline	✓ Meets	⚠ Partial	Diffit's 17 pages are paced with appropriate density for HS academic work. Gemini's 5 pages are denser and use ASCII flow art (▼ symbols) in place of rendered graphics
1.5	Answer-key consistency	✓ Meets	⚠ Partial	Diffit includes two answer-key sections (pp14–15 with explanations; pp16–17 with **source-quoted references** for each answer). Gemini includes a grading rubric for the formative summary task but **no answer key for the six analysis questions** in Part 3 — a teacher grades those by hand
2.1	Standards alignment	✓ Meets	✓ Meets	Both packets structure activities around RH.9-10.2's verbs (central ideas + how ideas develop). Diffit's matching tables and Gemini's "Step 1 / Step 2" analysis questions both directly probe the standard's demand
2.2	Cognitive depth	✓ Meets	✓ Meets	Diffit: recall (MC quiz), application (modern-day vocab examples), analysis (criticism/counter-argument matching), evaluation (synthesis reflection). Gemini: comprehension + analytical "development of ideas" questions + objective summary task
2.3	Genuine variety	✓ Meets	⚠ Partial	Diffit: 8 distinct activity types asking meaningfully different cognitive work. Gemini: primarily one type (analysis questions following excerpts) plus the summary task
2.4	Question quality	N/A	N/A	MC quality is N/A for Diffit on this scoring pass.
2.5	Honesty about what the standard requires	✓ Meets	✓ Meets	RH.9-10.2 demands "how key events or ideas develop over the course of the text" — both tools build activities that probe argument development, not just recall

#	Criterion	Diffit	Gemini	Evidence
3.1	Multi-activity coherence	✓ Meets	✓ Meets	Both packets reference common content throughout
3.2	Complete and ready to teach	✓ Meets	⚠ Partial	Diffit: passage + 8 activities + 2 answer keys, no placeholders. Gemini: passage + excerpts + 6 analysis questions with no answer key, 1 summary task with a grading rubric
3.3	Considered defaults	✓ Meets	✓ Meets	**Major change from the elementary case.** Gemini's first response was usable here — no follow-up needed. Diffit's default output (with optional follow-ups skipped) was also usable. Both tools handled HS-level prompting better than elementary
3.4	Honest layout for student work	✓ Meets	⚠ Partial	Diffit's tables and writing prompts have appropriate space per task. Gemini's analysis questions list under "Step 1" and "Step 2" with no dedicated writing space
3.5	Classroom-workflow format fidelity	✓ Meets	✓ Meets	No broken LaTeX or rendering issues in either artifact
4.1	Right reading level	✓ Meets	✓ Meets	Both pitched at HS academic register
4.2	Engaging voice	✓ Meets	✓ Meets	Both use narrative voice for context and primary-source quotation for content
4.3	Differentiation in the workflow	N/A	⚠ Partial	Tested in the Differentiation in practice section below (N/A here for Diffit).
4.4	Pedagogically-useful visuals	N/A	⚠ Partial	A text-only primary source does not require a conceptual diagram, so this is N/A for Diffit.
4.5	Asset-based framing	✓ Meets	✓ Meets	Both handle civil-rights content with appropriate seriousness

HS scorecard: Diffit 17 Meets, 0 Partials, 0 Fails (out of 17 applicable; §2.4, §4.3, §4.4 N/A for Diffit). Gemini 10 Meets, 7 Partials, 0 Fails (out of 18 applicable). N/A on §2.4 for Gemini.

What changed at HS, and what didn't

What changed: - **Gemini included a reading passage and primary-source excerpts.** Its elementary Tubman output didn't — students had no passage. The MLK packet has a 1-page narrative context and 1 page of substantive quoted excerpts from the letter, organized by theme (outside-agitator charge, direct action, just/unjust laws, white moderate critique). That's a real product behavior difference at HS. - **Gemini's first response was usable.** No follow-up was required to get a student-facing artifact. The §3.3 considered-defaults Fail from the elementary case does not replicate at HS. - **Gemini structured activities around the standard's verbs.** "Step 1: Determining Central Ideas" and "Step 2: Analyzing the

Development of Ideas" directly mirror RH.9-10.2's demand. - **Gemini included a grading rubric for the summary task.** Pedagogically sophisticated; not seen at elementary.

What didn't change: - Diffit still produces a more comprehensive packet — 17 vs 5 pages, 8 vs 2 activity types, two answer keys vs one rubric. - Diffit still includes activities (vocab table with modern examples, crossword, MC quiz with educative distractors) that Gemini's packet lacks. - Diffit still attributes all content to the single named source. Gemini drifted to other sources without being asked. - Diffit still surfaces follow-up refinement options without requiring them — the "Skip all" path produces a complete artifact on its own.

The honest read

At HS, **Gemini is closer to parity than at elementary, but Diffit still leads on completeness, activity variety, answer-key depth, and source discipline.** The pattern holds, just at narrower margins. A district admin who reads this and concludes "Gemini is good enough at high school but not at elementary" is reading the evidence correctly.

If anything, the HS comparison strengthens credibility — a side-by-side where Diffit wins every criterion at every grade band would read as a setup. The fact that Gemini's HS output is genuinely improved over its elementary output (and that we say so) is what makes the elementary findings land.

Visual evidence: where the divergence begins

The scoring patterns in the tables above are not a matter of one model being better than another. They are downstream of how the two products structure their input. The screenshots that follow show what a teacher actually sees and types in each tool — and the difference explains most of what the scoring tables found.

Diffit's input UI is structured

!Diffit's input UI for the Tubman ELA brief

Three things to notice:

1. **The source is a chip, not text.** "nps.gov" attaches as a structural input — Diffit knows it's the source and treats it accordingly. The teacher does not have to write a URL into a prose prompt and hope the tool ingests it correctly.
2. **Grade and standard are chips.** "Adapt: 5th Grade" and "Align: CCSS.ELA-Literac..." are structured selections from pickers, not phrases buried in a prompt sentence.
3. **Contextual suggestions at the bottom.** "Scaffold", "Vocabulary", "Bilingual", "Visual Supports", "Sentence Frames" are one-click affordances — they're the actions a teacher most often takes after

generating a packet, surfaced as buttons.

The same screen for the math brief shows the same pattern, plus a hint at what Diffit's UI does for teachers who don't know exactly what to ask for:

!Diffit's input UI for the math brief, showing prompt-completion suggestions

Below the input field, Diffit suggests four prompt completions inferred from the grade + standard context: "on decimal fractions", "covering multi-digit multiplication", "with collaborative problem-solving stations", "to support struggling learners." That's a teacher-friendly UX touch that does not exist in a chat interface — and one of the reasons Diffit's §3.3 (considered defaults) score is what it is.

Gemini's input UI is a chat box

!Gemini's chat compose for the ELA brief

Everything must be packed into one prose message: the standard code, the standard's text definition, the source URL. There is no structural acknowledgment that "this is the source" or "this is the standard" — they are just words in the message, indistinguishable from any other words. Gemini guesses what is what.

Same for math:

!Gemini's chat compose for the math brief

The teacher had to know to type "for 5th graders" *and* "that I can print" to give Gemini the framing it needed. Even so, Gemini's first response rendered every fraction as raw LaTeX (see the first-response artifact) — because in a general-purpose chat interface, LaTeX is a reasonable default for math notation. It is just not a reasonable default for a 5th-grade printable worksheet.

The follow-up moment

The §3.3 (considered defaults) Fail in the ELA scoring is grounded in this exact exchange:

!Gemini chat showing the teacher's follow-up: "Turn this into a student facing handout"

Gemini's first ELA response was a teacher-facing analysis, not a student-facing handout. The teacher had to type "Turn this into a student facing handout" before Gemini produced anything a student could use. That's a follow-up a savvy teacher might think to write — and one a less-experienced teacher might not, accepting Gemini's first response and walking into class with materials that don't fit the lesson.

Why this matters

Every score in the tables above traces back to one of these UI screens. Diffit's structured chips drive its standards-alignment, source-fidelity, and considered-defaults scores. Gemini's chat box drives its iteration burden, its LaTeX rendering default, and its tendency to interpret "packet" as "longer chat reply." The model underneath either tool is not the issue. The product surface is.

The five "look here" moments

If a district admin reads nothing else, these are the moments to look at.

1. Gemini's first math response was raw LaTeX

Open gemini-math-fractions-first-response.pdf, page 1, problem 1:

> "Lily spent **$\frac{1}{2}$** of an hour reading her book and **$\frac{1}{3}$** of an hour practicing the piano."

That's literally what the page renders. Every fraction in every problem, every entry in the answer key — raw LaTeX source. Gemini opens the response with *"here is a clean, ready-to-print worksheet"*; it isn't. A teacher who exports this to a Google Doc and hands it to a student gives them a math worksheet they cannot read.

The user had to follow up with "fix the rendering" to get a usable version. The corrected version is fine — but a teacher who didn't notice the issue, or assumed Gemini's first response was good, would walk into class with broken materials.

2. Gemini produced no reading passage

The ELA brief said "create a reading packet on Harriet Tubman, source = NPS." A reading packet, by name, contains something to read. gemini-ela-tubman.pdf has none.

What Gemini produced is an analysis worksheet — summaries of the NPS source organized into four "Interactions" sections with text-evidence callouts, followed by three open-ended questions. A student handed this packet has no continuous passage to read; they have a structured summary of a passage the teacher must obtain elsewhere.

Diffit's ELA packet opens with a 530-word original reading passage with five section headings, plus three contextual images, plus three activities, plus a 2-page answer key.

3. Diffit's math packet includes the visual fraction models the standard requires

CCSS.MATH.CONTENT.5.NF.A.2 explicitly requires students to solve word problems "by using visual fraction models or equations to represent the problem." Diffit's packet has a dedicated 4-page **Visual Fraction Models Practice** section on pages 1–4 with circle and bar models for each problem. Gemini's packet contains zero visuals.

A teacher who uses the Gemini worksheet to fulfill 5.NF.A.2 has, technically, not met the standard.

4. Diffit's error-analysis activity teaches against the textbook 5th-grade misconception

Pages 10–12 of Diffit's math packet include a section called **Reasonableness Check: Error Analysis**, featuring "Fraction Freddy" — a fictional student whose work shows three common misconceptions. The first is labeled *"The 'Add Across' Error"*:

> Freddy's Work: 1/2 + 1/3 = 2/5

This is the single most common 5th-grade fraction error in the curriculum — students who add both numerators and denominators. Diffit's packet doesn't just give students problems; it teaches them to diagnose and correct their own most likely mistake. That's a direct hit on Constitution §2.4 (educative distractors).

Gemini's packet has six word problems and an answer key. No error analysis, no misconception work.

5. Both ELA artifacts cite real facts from the source — but only one delivers a teacher-ready packet

This one matters for credibility: **Gemini did not fabricate.** Every quoted phrase in its ELA artifact appears on the NPS source page. Diffit's facts are likewise all source-verified.

So the difference between the two tools is *not* that one hallucinates and the other doesn't. The difference is that Diffit's source fidelity is delivered inside a complete classroom-ready packet (passage + activities + key + visuals), while Gemini's source fidelity is delivered inside a partial draft that's missing the pieces a teacher needs.

That's a more honest — and more useful — comparison than "AI hallucinates." It tells admins what they're actually buying: not error-prevention, but completeness, structure, and considered defaults.

What we don't claim

A few things this comparison is honest about — both because honesty is the right posture, and because over-claiming weakens everything else.

Gemini did not fabricate or hallucinate facts in either packet. Source fidelity for both tools was clean. Where Diffit and Gemini diverge is in completeness, structure, and classroom-readiness — not accuracy. We are deliberately not making a "Gemini hallucinates" argument because the evidence doesn't support it.

Two prompts is not a complete evaluation. This comparison covers ELA and math at one grade level with one prompt each. Other prompts will exercise different rubric criteria — translation faithfulness, collaborative-learning structures, source ingestion from a PDF rather than a URL, multi-grade differentiation, etc. We plan to publish those as separate comparisons.

Differentiation scoring is workflow-based. §4.3 scores Diffit Meets because Diffit offers post-generation reading-level adaptation and translation as workflow actions. This is documented product capability and was not exercised in the artifact runs themselves. A skeptical admin should verify this in a Diffit demo.

Gemini may improve over time. This comparison was run on Gemini for Education as of May 2026. Gemini's underlying model and education-specific features change. Re-running this comparison in 6 months is the right way to keep this honest. Methodology is documented below.

This is enablement, not academic research. The rubric is grounded in WestEd's framework and the Diffit Constitution, but the scoring was performed by a single evaluator, and the prompts were chosen to exercise the rubric well. A district considering both tools should run their own comparison on their own prompts before deciding.

What this comparison shows about the underlying products

Three structural differences explain the pattern of scores above. None of them are about model quality. They're about how the products are designed.

Diffit is a structured-input product. Gemini is a chat product. Diffit's UI gives a teacher dedicated fields for topic, grade, standard, and source — each is a typed input the system uses deliberately. Gemini's UI is a chat box; the teacher types prose and Gemini interprets. Structured input means Diffit can always know exactly what reading level, what standard, and what source the teacher meant. Gemini has to guess from the prose, and when it guesses wrong, the teacher iterates.

Diffit is built around the format of a teacher resource. Gemini is built around the format of a conversation. When Diffit's pipeline runs, it produces a multi-section packet: passage, activities, vocab, key. The product knows what a "packet" is. Gemini's pipeline runs as a chat turn — it produces whatever fits in a chat reply. When the teacher asks for a "packet," Gemini interprets that as a request for a long-form chat response and structures the response accordingly. That's why Gemini's ELA artifact has no passage: a packet, to a chat product, is just a longer chat reply.

Diffit's defaults are tuned for the classroom. Gemini's defaults are tuned for general utility. Gemini's first response to the math prompt was raw LaTeX because that's a fine default for general technical writing — math notation expressed in LaTeX is standard in academic contexts. It's not a fine default for a 5th-grade printable worksheet. Diffit's default rendering pipeline knows the output is meant for a 5th-grade student, so fractions render as visual fractions and answer keys export cleanly. The Constitution calls this

"considered defaults" — and it is the largest single quality differentiator between a purpose-built educational AI and a general-purpose one.

Methodology — reproducing this comparison

Anyone reading this should be able to re-run the comparison and verify the findings. Here's how.

Prompts

ELA prompt for Gemini:

> Create a 5th grade reading packet on Harriet Tubman. Align it to CCSS.ELA-LITERACY.RI.5.3 ("Explain the relationships or interactions between two or more individuals, events, ideas, or concepts in a historical, scientific, or technical text based on specific information in the text."). Use this source: https://www.nps.gov/hatu/learn/historyculture/htubman.htm

ELA inputs for Diffit: - Topic: Harriet Tubman - Grade: 5 - Standard: CCSS.ELA-LITERACY.RI.5.3 (selected from Standards picker) - Source: https://www.nps.gov/hatu/learn/historyculture/htubman.htm

Math prompt for Gemini:

> Create a 5th grade math practice worksheet on adding and subtracting fractions with unlike denominators (word problems). Align it to CCSS.MATH.CONTENT.5.NF.A.2.

Math inputs for Diffit: - Topic: adding and subtracting fractions with unlike denominators - Grade: 5 - Standard: CCSS.MATH.CONTENT.5.NF.A.2 (selected from Standards picker) - Source: none

HS ELA prompt for Gemini:

> Create a 9th–10th grade reading packet on Martin Luther King Jr.'s "Letter from Birmingham Jail." Align it to CCSS.ELA-LITERACY.RH.9-10.2 ("Determine the central ideas or information of a primary or secondary source; provide an accurate summary of how key events or ideas develop over the course of the text."). Use this source: https://kinginstitute.stanford.edu/encyclopedia/letter-birmingham-jail

HS ELA inputs for Diffit: - Topic: Letter from Birmingham Jail - Grade: 9th - Standard: CCSS.ELA-LITERACY.RH.9-10.2 (selected from Standards picker) - Source: https://kinginstitute.stanford.edu/encyclopedia/letter-birmingham-jail - Optional follow-up questions (engagement type, focus area): **skipped** to capture true default behavior

Conditions

- Diffit.
- Gemini for Education (Google Workspace for Education tier), May 2026.
- First responses captured before any follow-up prompting. Where follow-ups were needed to make Gemini's output usable, both versions captured and scored separately.
- All artifacts exported as PDF.
- Scoring against rubric.md, one evaluator pass with evidence cited per criterion.

Artifacts archived

All six artifacts are in artifacts/.

Citation

This comparison applies the Diffit Quality Constitution (drawing on WestEd's research into how teachers judge instructional-materials quality) to a side-by-side artifact comparison between Diffit and Gemini for Education.

Bugler, D., Marple, S., Burr, E., Chen-Gaddini, M., & Finkelstein, N. (2017). *How Teachers Judge the Quality of Instructional Materials: Selecting Instructional Materials, Brief 1 — Quality.* WestEd. https://www.wested.org/resource/selecting-instructional-materials-brief-1-quality/