

§22584(A)(4)(A)-(G) · THE PROCESS, PUBLISHED FOR REVIEW

Six Steps Between A Student's Words And A Model Release. Here Is Each One, Written For Your Review.

The statute says a school may review an operator's deidentification process before any use. This is that document. It describes what a response record carries inside the service, what is removed before a response can enter a reference corpus, how removal is verified, and how your school says no. Nothing here is summarized elsewhere in more detail; this is the process.

STEP 01

Eligibility

Only Partners Whose Agreement Covers It, And Never A School That Has Said No.

A school's responses are eligible to enter a reference corpus only when two things are true: the school's agreement covers deidentified use for model improvement (MSA §4(d)), and the school has not told us no. This page is how the process is made available for review before any use, as §22584(a)(4)(E) requires. A school does not have to review it (§22584(q)).

Eligibility is recorded per school, per release, in the corpus manifest described in Step 04. A school that tells us no, at any time, is left out of the next corpus and every one after it (Step 06).

[§22584\(a\)\(4\)\(E\), \(G\)](#) · [MSA §4\(b\), §4\(d\)](#)

STEP 02

Field removal

Everything That Is Not The Words Is Dropped.

Inside the service, a response record carries the fields a school needs to run the program: who the student is, which school and class, when they responded, what they were asked, and what a reviewer saw. Before a response can enter a reference corpus, every one of those fields is removed. What remains is two things: the text of what the student said, and the rubric level an expert rater assigned it. The record below is an example, with an invented student, so you can watch it happen.

RESPONSE RECORD		AS IT ENTERS THE REFERENCE CORPUS
student_name	Example Student (invented)	
student_id	PP-2031-0417	
service_token	a9f3...c21e (anonymous identifier used for scoring)	
school_lev	Example Charter Example County	
grade_class	8 Period 3 Advisory	
demographics	Gender, EL status, program flags as furnished by roster	
recorded_at	2026-10-14 09:42 PT	
device_modality	iPad spoken (audio 00:01:12)	
prompt_id	GRIT-03 "Tell me about something that took a long time to finish."	
media_uri	s3://.../responses/.../a9f3c21e.webm	
reviewer_released_at	Qualified reviewer 2026-10-14 15:10 PT	
words	"I wanted to quit robotics after we lost the first two matches, but I stayed after school for three weeks fixing the arm. It still wasn't perfect at the last competition, but it worked, and I figured out that the fixing part was the part I actually liked."	

rater_level

Grit · L3 "The Whole Arc" · expert rater, anchor set 6.0

2 fields enter the corpus · 11 removed

Struck fields are removed, not masked or hashed. There is no key, table, or salt that maps a corpus entry back to a record. The service_token is the anonymous identifier the scoring engine uses inside the service; it is a persistent identifier under §22584(a)(3), so it goes too. What enters the corpus cannot be linked to a person, a household, or a school, and the model cannot tell whether the words were spoken or typed, by whom, where, or when.

The prompt identifier is dropped as well. Rubric levels are defined per competency, not per prompt, so the corpus needs the competency name (Grit) and the level, not which of the competency's prompts elicited the response.

§22584(a)(4)(A), (D) · 34 C.F.R. §99.31(b)(1) · MSA §1(a), §4

STEP 03**Text redaction****Names And Places A Student Says Out Loud Are Removed From The Words Themselves.**

Field removal handles the record. It does not handle a student who says "my coach Mr. Alvarez at Pleasant Valley" in the middle of an answer. FERPA's deidentification standard asks whether a reasonable person in the school community could identify the student from the information, including what is in the text. So the words are reviewed before entry. The technical name for this is named-entity redaction with surrogate substitution: names and places are found automatically and replaced with a typed stand-in.

- An automated pass identifies personal names, school and district names, street addresses, phone numbers, email addresses, and place names below the level of a U.S. state.
- Each is replaced with a neutral placeholder of the same kind: [teacher], [school], [friend], [city]. The rubric scores what a student did and understood, not who was in the room, so the placeholder preserves what the rater needs.
- The expert rater who scores the response for the corpus reads it after replacement and flags any identifying reference the automated pass missed. A flagged response is corrected or excluded before entry. Raters are trained on this step and it is part of the rater agreement protocol.
- Responses that remain identifying after replacement, for example a detailed account of a widely reported local event, are excluded rather than edited further.

§22584(a)(4)(A), (D) · 34 C.F.R. §99.3 "personally identifiable information" (f), (g)

STEP 04

Manifest and test

Every Release Carries A Manifest, And The Corpus Is Tested Before It Is Trusted.

A corpus manifest is the chain-of-custody record for a training corpus. One is written for each model release and retained with the release in the model registry. It records, without identifying any student:

- Which partner authorizations the corpus draws on, by agreement reference, and each partner's review status under Step 01. The corpus itself carries no partner label; the manifest is the only place authorization is tracked, and it is not an input to training.
- The version of the field-removal rules (Step 02) and in-text review rules (Step 03) applied, and the count of responses excluded at Step 03.
- Response counts per competency and per rubric level, with no cell reported below ten responses.
- The result of the reidentification test below.

A specimen manifest is shown below so a reviewer can see the document itself, not a description of it. Before a corpus is used, IMPACTER attempts to reidentify it, which §22584(a)(4)(B)(ii) permits for exactly this purpose. Analysts who did not build the corpus attempt to link a sample of entries to service records using every field the service holds. The test passes only when no entry can be linked to a person, household, or school. A failed test returns the corpus to Step 02 or Step 03 and the release does not proceed. The pass or fail, the sample size, and the method are recorded in the manifest.

CORPUS MANIFEST **CM-7.0-01** **SPECIMEN** RELEASE V7.0 · IMPACTER MODEL REGISTRY

IDENTITY

manifest_id

CM-7.0-01

release

DeBERTa-v3-base + CORAL · v7.0 candidate

supersedes

CM-6.0-02

corpus_frozen

2026-11-03

corpus_digest

sha256 · 3e1f...9a4c (words + level file, post Step 03)

status

Candidate · awaiting Step 05 promotion gate

RULES APPLIED

field_removal_rules

FR-3.1 · 11 of 13 record fields dropped (Step 02)

in_text_review_rules

TR-2.0 · names, schools, addresses, contacts, sub-state places → typed placeholders (Step 03)

rater_protocol

RP-6 · flag-after-replacement, dual read on 10% sample

cell_floor

10 · no count below ten is reported

AUTHORIZATIONS DRAWN ON · BY AGREEMENT REFERENCE, NEVER BY SCHOOL NAME IN THE CORPUS

AGREEMENT REF.	AUTHORIZATION CLAUSE	REVIEW STATUS (STEP 01)	RESPONSES CONTRIBUTED	EXCLUDED AT STEP 03
MSA-2024-0117	MSA §4(d) · covered	Reviewed · found sufficient 2026-09-30	4,120	37
MSA-2024-0142	MSA §4(d) · covered	Not reviewed · §22584(q) · no objection	2,865	19
MSA-2025-0031	MSA §4(d) · covered	Reviewed · found sufficient 2026-10-14	1,940	12
MSA-2025-0088	MSA §4(d) · school said no	Prohibited · excluded by manifest	0	—

AGREEMENT REF.	AUTHORIZATION CLAUSE	REVIEW STATUS (STEP 01)	RESPONSES CONTRIBUTED	EXCLUDED AT STEP 03
Total		3 contributing · 1 excluded	8,925	68

COMPOSITION · RESPONSES PER COMPETENCY AND RUBRIC LEVEL					
COMPETENCY	L1	L2	L3	L4	TOTAL
Grit	412	1,038	1,215	396	3,061
Curiosity	388	1,102	1,146	301	2,937
Self-Control	451	1,090	1,061	325	2,927
Total	1,251	3,230	3,422	1,022	8,925

REIDENTIFICATION TEST · §22584(A)(4)(B)(II)
conducted_by
Two analysts not involved in corpus construction
sample
500 corpus entries, stratified by competency and level
method
Attempted linkage to service records using every retained service field; text search on placeholders and residual proper nouns
result
0 of 500 linked to a person, household, or school · PASS
residual_findings
2 entries with an unreplaced place name below state level → corrected, TR-2.0 rule added, corpus re-frozen
completed
2026-11-03

HUMAN STANDARD · STEP 05 INPUTS
raters
6 expert raters, trained on RP-6; each response scored by one rater, 10% dual-scored
inter_annotator_agreement

Reported per competency in the v7.0 Model Card on release; promotion requires the published QWK threshold

held_out_split

15% of corpus, stratified, sealed before training

Specimen manifest with illustrative values. The manifest for the release your school uses is furnished on request (Step 06).

prepared · ML lead · reviewed · privacy officer · retained with release in registry

[§22584\(a\)\(4\)\(A\), \(B\)\(ii\)](#) · [MLflow Model Registry Lineage](#)

STEP 05

Release governance

A Release Is A Governed Event, Run Like A Research Protocol, With A Versioned Result.

The comparison a school will recognize is an institutional review process. Participants are enrolled by written authorization (Step 01). The protocol is fixed and reviewable in advance (Steps 02 through 04). Human judgment sets the standard: expert raters score the deidentified corpus against the published rubric, and their inter-annotator agreement is measured and published before the corpus is trusted. The model is trained on that human-rated standard and evaluated on held-out responses against the same raters. It is promoted only if it clears the published agreement threshold, then frozen, versioned, and documented in a Model Card that states its training, evaluation, fairness testing, intended uses, and out-of-scope uses.

A frozen release does not learn in production. Scoring your students' responses returns a level; it changes nothing in the model. The release your school uses is the release described in its Model Card, and the manifest for that release is available to you.

[§22584\(c\)](#) · [Model Card v6.0](#) · [Inter-Annotator Agreement](#) · [Technical Report 2025-01](#)

STEP 06

Your review and your veto

Everything Is On This Page. Tell Us No, And Your Data Is Out.

ALREADY HERE, FOR YOUR REVIEW

- The process, Steps 01 through 05 above
- Every field removed before use (Step 02)
- The rules for taking names and places out of students' words (Step 03)
- A specimen manifest and the reidentification test (Step 04)
- Every subprocessor, and what each one is bound to
- [Our written data retention policy](#)
- [The Master Subscription Agreement](#)

This is the review §22584(a)(4)(E) asks for. Your school does not need to request it, and does not need to review it (§22584(q)).

IF YOUR SCHOOL SAYS NO

Tell us in writing that your students' data may not be used, even deidentified. That is your right under §22584(a)(4)(G) and MSA §4(d). We remove your students' data from the training corpus, keep it out of every future release, record the exclusion in each manifest, and confirm in writing.

Your own standard. If your school keeps a deidentification policy, send it and we apply it (§22584(a)(4)(F)).

If you do nothing. Your students' data may be used only in deidentified form, only as this page describes, and never to train a generative model. You can say no at any time.

[§22584\(a\)\(4\)\(E\), \(F\), \(G\), \(q\)](#) · [MSA §4\(d\)](#)
