NOVARA
CONSULTING GROUP

Scores Without Footnotes: AI-Written Evaluation Reports and the Deaf Child

Heather Grizzle · November 12, 2026

AI report writers turn score files into evaluation reports. For a deaf child tested through an interpreter, the numbers arrive without the footnotes that give them meaning.

Contents

1. The Reevaluation
2. What the Report Writer Sees
3. Tests That Were Never Built for Her
4. Eli's Evaluation and the Overlap Nobody Scores
5. What Jordan's Teachers Saw
6. The Numbers Lose Their Footnotes
7. The Shortage Underneath
8. What an Evaluation in Her Language Looks Like
9. The Contracted Evaluator and the Policy Gap
10. What Jordan's Mother Read
11. Two Machines, One Child
12. Should the Report Say a Machine Drafted It?
13. A Field the Report Writer Cannot Skip
14. The Cost of Doing It Right
15. The Evidence Before the Decision
16. Questions for Governance Review

ABOUT THIS PIECE

This is the ninth part of NCG's AI x IEP series. Part seven read the special factors sections of six deaf and hard of hearing students at Pine Lake and found one generated sentence in all of them.[1] Part eight went back to the screeners that decide which children get looked at.[2] This part follows the next step, the evaluation, and the software that now writes the report a team reads when it decides what a child's disability is and what she needs. It ends, as each part now does, with questions for governance review. Pine Lake, Jordan, her family and the evaluator are fictional composites, and the report language quoted is illustrative. The products' published claims, the research, the professional standards, the federal rules and the National Deaf Education Data Index figures are real.

IN BRIEF

- Evaluation reports are the evidence an IEP team uses to decide eligibility and needs. AI report writers that build those reports from uploaded score files are now sold to school psychologists, and a 2024 national survey found about two-thirds of school psychologists had used AI in the past six months, most commonly to generate recommendations.
- A score file carries numbers. It usually does not carry the conditions under which a deaf child earned them: an interpreter between examiner and child, a test normed on hearing children, a narrowing visual field, a session stopped early for fatigue. Those conditions are what make the numbers mean something.
- Professional testing standards call for departures from standard administration to be documented and reported. Federal rules require that assessments be used for purposes for which they are valid and reliable for the child. A report that turns interpreter-mediated scores into standard interpretive prose can satisfy neither while reading as if it satisfies both.
- The shortage underneath is human. Fewer than one percent of speech-language pathologists are qualified to provide services in American Sign Language, by one published estimate. Where no fluent evaluator exists, the software inherits the gap and writes over it.
- The fix is a required field: how the test was given, in what language, with what departures, carried into every sentence that interprets a score. Some states already offer evaluation by fluent signers. Districts can ask for both.
- Contracted evaluators often work outside any district's AI policy, and families are rarely told when a report was drafted by software. Both are fixable in a contract and a sentence.
- The piece closes with sixteen questions for governance review.

The Reevaluation

Jordan, the fourteen-year-old from part seven who was born deaf and is losing her peripheral vision to Usher syndrome, was due for her three-year reevaluation in the winter. Federal rules require a district to reevaluate every child with an IEP at least once



every three years unless the parent and district agree it is unnecessary, and Jordan's mother did not agree.[3] Her daughter's needs had changed more in three years than in the nine before them. She wanted to know where Jordan stood, and she wanted the team to know it in writing.

Pine Lake does not employ a school psychologist who signs. Few districts its size do. It contracts for evaluations of deaf and hard of hearing students with a psychologist who covers four districts, a careful and experienced evaluator who has tested deaf children for years through interpreters, and who knows the limits of that arrangement better than anyone at Pine Lake. He spent three mornings with Jordan in January. He gave a standard cognitive battery, an academic achievement battery and a set of rating scales, with an educational interpreter from Jordan's school in the room. He made notes on the protocols as he went, the way good evaluators do: the verbal subtests were interpreted and their norms do not apply; Jordan missed two items that were signed at the edge of the interpreter's signing space and passed both when he repeated them closer; on the visual subtests she scanned the page slowly from edge to edge, which he attributed to her field loss rather than her processing; the third session ended early because her eyes were tired.

It helps to understand the arithmetic of his week, because it explains everything that follows. A contracted evaluator who covers four districts is paid by the evaluation, drives between buildings, tests in whatever room is free, and writes at night. A comprehensive report for a child with a complicated profile can take a full day to write well, and a three-year reevaluation for a deaf child who is losing her vision is about as complicated as profiles come. Every hour he spends writing is an hour he is not testing the next child on a waiting list that, in his districts, runs several weeks deep. The pressure on him is not to cut corners. It is to finish, and the software was sold to him as a way to finish.

Then, back at his office with seven reports due that week, he did what about two-thirds of his colleagues now report doing in some form.[4] He uploaded the score reports to an AI report writer.

What the Report Writer Sees

The category is new and real. One major test publisher, PAR, launched an AI Report Writer in late 2025 whose first template supports psychoeducational evaluations, IEP

planning and IDEA eligibility for children from preschool through grade twelve. The company says it automatically pulls scores and key details from uploaded files, produces a fully customizable draft, and saves frequent users an average of six hours a week.[5] Its published answers to common questions are, on the points they address, careful: uploaded data is used only for that session and not retained, it is never used to train a model, the system is closed and hosted privately, and the psychologist remains responsible for reviewing, editing and adding professional judgment. The answers also note that some test results are unsupported because they have not been validated for AI-assisted use.[6] NCG found nothing in those answers about tests administered through an interpreter or in sign language, or about examinees who are deaf, hard of hearing or deafblind.[6]

That silence is not unusual and is not a sign of carelessness. It reflects what the tool works from. A report writer of this kind reads what the evaluator uploads, and the heart of that is the score report: the test names, the dates, the standard scores, the percentiles, the confidence intervals. It can also take background, interview and observation notes, if the evaluator writes them up and uploads them. Jordan's evaluator uploaded the score reports and the referral background. The score report the scoring software printed was clean. It did not say that an interpreter had rendered the verbal subtests into ASL. It did not say that two items were repeated. It did not say that the third session ended early. Those facts lived in the psychologist's handwriting on the protocols, which test publishers rightly do not want uploaded anywhere, and in his memory.

From the score report, the generator wrote what such generators write. Jordan's verbal comprehension fell in the low average range. Her visual spatial and processing speed scores fell below average, "suggesting relative weaknesses in visual processing and speed of information processing that may impact her ability to complete tasks efficiently." Her academic achievement was below grade level in reading comprehension and written expression. It recommended extended time, preferential seating, chunked assignments and graphic organizers.



What reaches the report writer and what does not. The score file carries the numbers; the conditions that determine what the numbers mean stay on the protocol and in the evaluator's memory. Novara Consulting Group; illustrative of the fictional Pine Lake evaluation

Every sentence was a fair reading of the numbers on the page. Several were wrong about Jordan. The processing speed and visual spatial subtests depend on scanning visual material quickly, and a child with a narrowing visual field scans slowly for a reason that has nothing to do with how fast she processes information. The verbal comprehension score came from subtests normed on hearing children who took them in spoken English, administered to a deaf child through an interpreter, which changes what is being measured in ways no norm table accounts for. And the recommendation of preferential seating, for a student who needs an interpreter positioned in her remaining field of vision against a plain background in good light, is the same average-of-the-category answer part seven found in her special factors.[1]

Tests That Were Never Built for Her

The software's blind spot sits on top of an older one. Most of the instruments in a school psychologist's kit were normed on samples of hearing children who took them in spoken English, and their norm tables say how a child's performance compares with those children under those conditions. A deaf child who takes the same test through an interpreter is not taking the same test. The interpreter has to choose how to render each instruction and each item, some items depend on English sound or vo-

cabulary in ways that have no clean ASL equivalent, and the child's answers come back to the examiner through a second translation. The result can be a useful description of how the child performed in that session. It is not a standard score in the sense the norm table means, and a careful report says so.

Evaluators know this, and the common workaround has its own problems. Many lean on nonverbal measures for deaf children, on the reasoning that a test that does not require spoken language is fairer. That is often the right call, and it still leaves gaps. Nonverbal tests still have instructions that must be understood, still depend on vision, and still compare the child with hearing peers. For a child like Jordan, whose visual field is narrowing, a test built to avoid her hearing loss runs straight into her vision loss. The federal rules require a child to be assessed in all areas related to the suspected disability, including vision and hearing, and the evaluation to be comprehensive enough to identify all of her needs.[7] A cognitive profile built from visual subtests, without a functional vision assessment, can attribute to her mind what belongs to her eyes.

The rating scales have the same problem from a different direction. Teachers and parents complete behavior and social-emotional scales by rating how often a child does things: asks for help, joins in, responds when spoken to, follows directions given to the group. A deaf child in a hearing classroom may do all of those things rarely, not because of anything a behavior scale is designed to measure, but because she cannot follow the conversation those behaviors depend on. A teacher who does not sign, rating a child she cannot easily talk to, is rating the communication barrier as much as the child. The scale's scoring software does not know that. Neither does a report writer that reads its output.

None of this is new to the professionals who evaluate deaf children well, and it is the reason the best of them write long, cautious reports that spend as many paragraphs on conditions as on scores. What is new is a tool whose input is the scores, whose output is fluent interpretation, and whose promise is that the evaluator will spend less time writing. The caution was the part that took the time.

Eli's Evaluation and the Overlap Nobody Scores

Jordan's case is about a second disability hiding inside the scores. Eli's, from part seven, is about two disabilities that look like each other.[1] Eli is Deaf and autistic, and his

evaluations over the years have had to separate what comes from each. That is one of the hardest problems in the evaluation of deaf children, and it is one an AI report writer is especially poorly placed to handle.

Many of the behaviors that autism rating scales and observation tools ask about are behaviors that depend on communication: whether a child responds to her name, joins in conversation, uses gestures and language to share attention, understands what others are saying, plays with peers. A deaf child who has had poor access to language, or who is surrounded by people who do not sign, can show some of those patterns for reasons that have nothing to do with autism. A deaf child who is autistic can have the autism missed or delayed because the same patterns are attributed to deafness. Evaluators who work with deaf children know this overlap well, and the careful ones spend a large share of their reports explaining how they separated the two: which behaviors appeared in signing environments as well as hearing ones, what the child's teacher of the deaf observed, what the family sees at home.

None of that reasoning is in a score report. A rating scale completed by a teacher who does not sign produces a set of elevated scores in social communication, and a report writer working from those scores will describe a profile consistent with autism, which for Eli happens to be correct and for another deaf child in the same classroom might not be. The tool cannot tell the difference, because the difference lies in the conditions under which the ratings were made and in the evaluator's judgment about them. When an evaluation turns on separating two explanations for the same behavior, the most important paragraphs in the report are the ones that do the separating, and they are the paragraphs a tool built to interpret scores will not write.

Eli's most recent evaluation was done well, by a psychologist who had worked with deaf autistic students for years and who wrote every word of the report herself. His team mentioned that, in passing, as the reason his plan had been right for so long. It is also the reason his file was the only one of the six in part seven whose special factors section had been written by people before the generator arrived.

What Jordan's Teachers Saw

A formal battery is not the only evidence an evaluation can draw on, and for a child like Jordan it may not be the best. Federal rules require the team, as part of any reevaluation, to review existing data on the child, including classroom-based observations and observations by teachers and related services providers, along with informa-

tion from the parents.[8] For a deaf child tested through an interpreter, those observations are often more valid than the scores, because they were made in the child's actual environment, over months, by people who can communicate with her.

Jordan's teacher of the deaf had two years of observations: how she followed signed instruction when the interpreter stood in the right place and how she lost it when he did not, how her reading of signs at the edge of her field had changed, how she tired after long visual tasks, and how her written English had grown when she was taught through ASL. Her mother had a decade. Neither source was in the score report, and so neither reached the report writer. The revised report drew heavily on both, and its most useful pages, by her team's account, were the ones that described what Jordan does in her classes and why, rather than the ones that ranked her against children who had taken the tests in spoken English.

An AI report writer could be built to take observations seriously. It could ask for them, structure them and carry them into the narrative alongside the scores, flagged as observations rather than measurements. Most are built, as their own descriptions say, to pull scores and key details from uploaded files.[5] That design choice tells an evaluator, every time he opens the tool, which kind of evidence it treats as primary. For deaf children tested through interpreters, it is usually the wrong kind.

The Numbers Lose Their Footnotes

Testing has rules for this situation, and they predate any software. The Standards for Educational and Psychological Testing, published jointly by the American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education, call for changes or disruptions to standardized administration procedures to be documented and reported to the people who use the results.[9] The federal special education regulations require that assessments be used for purposes for which they are valid and reliable, be administered by trained and knowledgeable personnel, and be given in accordance with the producer's instructions, and they require, as part eight described, that an assessment given to a child with impaired sensory skills reflect her aptitude rather than her impairment.[10] An evaluation team deciding eligibility and needs must draw on information from a variety of sources, and the parent must receive a copy of the evaluation report.[11]

None of those rules forbids an evaluator from using software to draft a report. What they require is that the report carry the conditions under which the scores were obtained, because the conditions are what tell a reader whether the scores can be interpreted the standard way. A careful evaluator writes those conditions into the report by hand. Jordan's evaluator would have, in a slower week. The risk with an AI report writer is not that it invents conditions. It is that it never learns them, produces fluent standard interpretations of numbers that came with footnotes, and hands a busy evaluator a draft in which the footnotes are already gone. To put them back, he has to remember them, find them on the protocols, and rewrite paragraphs that read as finished. Part four described the research on what busy people do with a finished-looking draft from an automated system.[12] Evaluators are not exempt from it.

The national picture suggests this is not hypothetical. In a survey of 199 school psychologists across the country in 2024, published in School Psychology in 2026, about two-thirds reported using AI in the previous six months, and roughly a quarter used it weekly or more. The most common use was generating recommendations, reported by 52.8 percent, followed by report writing. Ninety-four percent said they edited AI-generated content before using it. Only five percent said their workplace had an AI policy, most had no formal training, and about six percent had entered identifiable student information into an AI tool.[4] Those figures describe a profession adopting tools faster than its institutions are writing rules for them, which is the pattern this series has found at every stage. They do not separate out evaluations of deaf or hard of hearing children, and no study NCG could find has.

SCHOOL PSYCHOLOGISTS AND AI, 2024

2 in 3

used AI in the past six months

52.8%
used it to generate recommendations

94%
said they edit AI content before using it

5%
reported a workplace AI policy

Source: Farmer, Lockwood, Floyd and Sisco, School Psychology 41, no. 2 (2026); n = 199

The Shortage Underneath

It would be easy to make Jordan's evaluator the problem in this story, and it would be wrong. He did what the system around him allowed. Pine Lake has no psychologist who signs because there are very few to hire, and the same is true in most of the related professions that evaluate deaf children.

The scale of the gap is easiest to see in speech-language pathology, where it has been measured. A study by David Smith, Kristen Secora, Brittany Lee and Marie Coppola, published by the American Speech-Language-Hearing Association, cites estimates that about 42 percent of speech-language pathologists regularly serve deaf and hard of hearing children and that fewer than one percent of all speech-language pathologists are qualified to provide services in ASL. Surveying 107 clinicians against expert-recommended qualifications for working with signing deaf children, the authors found

that those who specialize in deaf and hard of hearing children met 81.3 percent of the standards, while generalists met 31.5 percent, falling shortest on the standards that concern ASL fluency.[13] Most deaf children who sign will be evaluated, at some point, by a clinician who does not.

The National Deaf Education Data Index shows the same shortage from the state side. Delaware has one educational audiologist statewide. Hawaii's department of education had fifteen positions for educational interpreters at the end of 2024 and three were filled, while an estimated 81 students needed interpreting.[14] Where there is no fluent evaluator, the evaluation runs through an interpreter, and where it runs through an interpreter, the scores carry footnotes that a report writer working from score files will not see.

Some states have built a way around the shortage. Washington's Center for Deaf and Hard of Hearing Youth sends an outreach team to districts across the state, including school psychology services by specialists with fluent signing skills and an ASL specialist who evaluates a child's receptive and expressive ASL.[15] A district that can call on a team like that does not have to choose between an evaluation in the child's language and an evaluation at all. Most states have nothing comparable, and in those states the honest description of many deaf children's evaluations is that they were conducted in a language the child does not use, with standard interpretation applied anyway.

The shortage is not only a staffing problem. It is a data problem, and AI tools inherit it as one. When most evaluations of deaf children run through interpreters, most of the evaluation reports that exist, the reports any model trained on evaluation reports would have learned from, describe deaf children through interpreted scores read against hearing norms. A tool trained on that record will reproduce its habits fluently. The scarcity of fluent evaluators shapes not just who tests a deaf child today but what the software that helps write her report has learned a deaf child's report looks like.

What an Evaluation in Her Language Looks Like

It is worth describing what the alternative looks like, because it is not exotic. An evaluation of a deaf child who signs, done well, starts with her language. Instruments built for that purpose exist. Gallaudet University's VL2 center publishes the Visual Communication and Sign Language Checklist, a standardized observational tool for tracking the development of signing children from birth to age five, and schools for the deaf



use a range of ASL measures, receptive and expressive, to document where a child's language actually stands.[16] A fluent evaluator can assess a child directly in ASL, observe her in a signing environment as well as a hearing one, and ask Deaf adults who know her, a teacher of the deaf or a mentor, how she communicates when she is not working through an interpreter.

For a child who is also losing her vision, a good evaluation adds what a cognitive battery cannot provide: a functional vision assessment of how she uses the sight she has, a learning media assessment of whether print, large print, braille or tactile formats will serve her as her field narrows, and an orientation and mobility evaluation. Every state has a federally funded deaf-blind project that can advise a district on all three.[17] The federal rules already require that the evaluation cover vision and hearing where they are related to the suspected disability.[7] For Jordan they are the disability.

None of that is cheap, and in many districts none of it is local. It is still the evaluation the regulations describe: one administered in the child's language and mode, by trained and knowledgeable personnel, using measures valid for the purpose, and comprehensive enough to find all of her needs.[10] An AI report writer cannot supply any of those conditions. What it can do, if it is designed carefully, is refuse to pretend they were met.

The Contracted Evaluator and the Policy Gap

The national survey found that only five percent of school psychologists reported a workplace AI policy.[4] For contracted evaluators like Jordan's, the gap is wider, because a contractor works under the policies of several districts at once, or of none. Pine Lake had no AI policy for evaluations when Jordan was tested. Its contract with the evaluator said nothing about software, data or drafts. The evaluator's own practice, a sensible one, was to use only a tool from a major publisher that promised not to retain or train on his data, and to remove names before uploading. That was his choice, not the district's requirement, and the next contractor might choose differently. About six percent of the surveyed school psychologists reported having entered identifiable student information into an AI tool.[4]

A district's contract with an evaluator is the cheapest place to close that gap. It can say whether AI drafting tools may be used for reports, which tools, and on what terms: no identifiable data beyond what the tool's privacy commitments cover, no uploading of

test protocols, no use of tools that retain or train on the data. It can require the evaluator to complete the administration conditions section described below before any interpretation is drafted. It can require the evaluator to keep the generated first draft with the signed report and produce both on request, for the reasons part four gave for IEP drafts.[12] And it can require a sentence in every report saying whether software was used to draft any part of it. None of those terms costs the evaluator much. Each of them turns a private practice into a district standard that a family, a monitor or a hearing officer can check.

What Jordan's Mother Read

Jordan's mother received the evaluation report a week before the meeting, as she had asked to. She read it the way part five's parent read the IEP draft: looking for her daughter in it.[18] She found the low average verbal comprehension score and the below average visual processing finding, and she found no mention of an interpreter, of norms, of the edge of the signing space, or of the fact that Jordan is losing her sight. The report referred to a hearing impairment once, in the reason for referral, and never again.

She called the evaluator. To his credit, he did not defend the draft. He had meant to add a section on administration conditions and had run out of week. He revised the report: a page describing how each test was administered and through whom, a statement that the verbal scores could not be interpreted against the norms and should be read as a description of Jordan's performance in interpreted English rather than her verbal ability, a note that the visual subtests were affected by her field loss and should not be read as a processing weakness, and a recommendation for an assessment of her ASL by a fluent evaluator and a functional vision assessment before any conclusion about her cognitive profile. His recommendations section changed from extended time and preferential seating to interpreter positioning, lighting, contrast, rest breaks for visual fatigue, and a referral to the state deaf-blind project.

The revised report was a better document and a less confident one. That is the right direction. A report that says what it does not know about a deaf child is more useful to her team than one that sounds complete. Jordan's mother also asked, in writing, for the independent ASL language assessment the revised report recommended, which put Pine Lake where part six left it with Nora: pay for an evaluation in the child's language, or explain to a hearing officer why it should not.[19] Dana paid.

It is worth noticing what made the difference, because none of it was unusual. Jordan's mother asked for the report before the meeting, which any parent can do. She read it looking for her daughter, and when she did not find her, she asked the person who wrote it to explain. The federal records rules give a parent the right to a response to reasonable requests for explanations and interpretations of her child's records, and an evaluation report is the record a parent most often needs explained.[20] Most parents do not know they can ask, and most evaluators are never asked. When they are, the good ones, like Jordan's, often welcome the chance to put back what the week took out.

Two Machines, One Child

> A report that says what it does not know about a deaf child is more useful to her team than one that sounds complete.

There is a second reason the evaluation report matters more than any other document in this series, and it brings the story back to where it began. Part one described what Pine Lake's IEP generator was given to write Nora's present levels: last year's IEP, a benchmark, writing samples and progress notes.[21] Next year, when Jordan's annual review comes around, the generator will be given her evaluation report too. If that report is the first draft, with the low average verbal comprehension and the visual processing weakness written as findings, the IEP generator will read them as findings. It will write a present levels statement that describes a slow visual processor with low average verbal ability, and goals to address both, in fresh and correctly dated sentences. One machine's unexamined draft becomes the next machine's evidence.

That is how a finding that should never have been made survives for years, and it is what part four found with Nora's language delay, carried for three years by people and then by a generator.[12] The audit in part four could trace it because the drafts were kept. An evaluation report drafted by AI deserves the same treatment as an IEP drafted by AI. Part five argued that a parent may ask for the records a proposal was built from, including the machine's draft.[18] An evaluation report is such a record, and its AI-generated first draft, if one exists and the district or its contractor keeps it, is the only document that shows which interpretive sentences the evaluator wrote and

which he accepted. A district contracting for evaluations can require its evaluators to keep the generated draft with the signed report, for the same reasons part four gave for keeping IEP drafts.

The repair to Jordan's report did one more thing that matters here. Because the revised report now opens with a section on how each test was given and what the scores can and cannot mean, the next machine to read it will read that section too. Whether it will understand it is a question for the vendor. Whether a person on Jordan's team reads it before pressing the button is a question for the district.

Should the Report Say a Machine Drafted It?

Jordan's mother did not know, when she first read the report, that software had drafted it. She found out only because she called the evaluator and he told her. That raises a question the profession has not settled.

The training offered to school psychologists on AI report writing treats disclosure to families and teams as a matter of judgment. One widely advertised 2026 webinar for school psychologists sorted AI uses into do's, don'ts and "depends," and placed interpretation, clinical judgment calls and disclosure to families and teams in the "depends" category.[22] That is an honest description of where the field is. It is not a satisfying answer for a family reading a document that will decide what services their child receives.

NCG's view is that an evaluation report drafted in part by software should say so, briefly and without apology: which sections were drafted with an AI tool, and that the evaluator reviewed and revised them. The reason is not that AI drafting is improper. It is that the report is evidence, and the people relying on it, the team, the family and, if it comes to that, a hearing officer, are entitled to know how the evidence was produced, in the same way the testing standards entitle them to know how each test was administered.[7] A sentence of disclosure also changes the evaluator's own relationship to the draft. An evaluator who must write "portions of this report were drafted with an AI tool and reviewed by the examiner" is more likely to have reviewed them, because he has put his name to the claim.

A Field the Report Writer Cannot Skip

The repair for this is narrow and almost free, which is why it is worth asking for. Every AI report writer used for an evaluation should require, before it drafts a single interpretive sentence, a structured field for administration conditions: the language and mode in which each test was administered, whether an interpreter was used and the interpreter's qualification, any departures from standard procedure, any sensory, physical or fatigue factors that may have affected performance, and whether the norms are considered applicable. When that field records an interpreter, a sign language administration, or a sensory factor, the tool should not produce standard norm-referenced interpretive language for the affected tests. It should produce a placeholder that says the evaluator must describe what the scores mean for this child, and it should carry the conditions into the report automatically.

None of that requires a test publisher to upload a protocol or a district to share anything more than it already does. It requires a few form fields and a rule. A district buying or permitting a report writer can ask its vendor for them, and a state department of education approving evaluation practices can require them. Part two proposed that districts write the review of an AI-drafted IEP into the vendor contract.[23] The evaluation report is upstream of the IEP. If the evidence the team relies on has already lost its footnotes, the most careful review of the IEP will not find them.

States have a role here that districts cannot play alone. A state that publishes guidance on evaluating deaf and hard of hearing students, or that maintains an approved list of tools for any purpose, can make the administration conditions field a condition of approval for any report writing tool used in its schools. A single state requirement would reach every contracted evaluator who works in that state, including the ones who, like Jordan's, cover several districts and answer to none of their policies in particular.

Administration conditions a report writer should require before it interprets a score

Field	Why it matters for a deaf or hard of hearing child	
Language and mode of administration	Scores from interpreted administration do not mean what the norms say	
Interpreter used and qualification	The interpreter is part of the test; a weak interpretation lowers the score	
Departures from standard procedure	Repeated items and modified instructions change what was measured	
Sensory and fatigue factors	A narrowing visual field or listening fatigue can look like a processing weakness	
Norms applicable	If not	the report must describe performance instead of ranking it

Source: NCG; drawn from the Standards for Educational and Psychological Testing (2014) and 34 CFR 300.304

The Cost of Doing It Right

The obvious objection to everything in this piece is cost, and it deserves a straight answer. An evaluation by a psychologist fluent in ASL, or by a team like Washington's, usually costs a district more than its standard contracted evaluation, may require travel or a remote session, and may take longer to schedule. A functional vision assessment, a learning media assessment and an orientation and mobility evaluation for a student like Jordan add more. Pine Lake's independent ASL evaluation for Nora, or-

dered by a hearing officer in part six, was more expensive than any evaluation the district had paid for that year.[19]

The comparison that matters is not between two evaluations. It is between one good evaluation and three years of services built on a bad one. Nora′s language delay finding, carried for three years, sent her to remediation she did not need and kept her from the bilingual instruction she did, and it ended in a due process hearing that cost Pine Lake far more in staff time, legal fees and the evaluation it was ordered to fund than an ASL evaluation at the start would have. Jordan′s first draft, had it been signed, would have sent her into processing-speed interventions for a weakness she does not have, while the changes she needs, in lighting, interpreter positioning, rest for visual fatigue and preparation for braille, waited. An evaluation is the cheapest point in the whole process at which to be right. Every later step inherits its mistakes.

AI report writers are sold on time saved, and the time is real. The question a district should ask is what the evaluator does with it. If the six hours a week a report writer saves are spent on more evaluations at the same depth, the district has gained capacity. If they are spent because the depth was removed, with the administration conditions, the observations and the caution left out of the draft and never put back, the district has bought speed at the cost of the only part of the report that made it evidence.

The Evidence Before the Decision

Nine parts in, the series has followed AI through the paperwork of special education from the back to the front: the IEP draft, the contract, the vendor terms, the audit, the records request, the hearing, the special factors, the screener, and now the evaluation report. At each step the software did roughly what it was designed to do, and at each step it did it on information that left out the one fact about a deaf child that mattered most: what language she uses, what she can see and hear, and what that means for every number written about her.

The evaluation report is where that omission does the most damage, because everything after it is built on it. An IEP team reads the report, finds a processing weakness, writes a goal for it, and measures progress against it for three years. Jordan′s revised report will spare her that. It took a parent who read the report before the meeting, an evaluator who admitted what he had left out, and a director who paid for an evalua-

tion in the child's language. None of those should have to depend on luck. The record is not the child, a score is not a finding, and a number that has lost its footnote is not evidence about a deaf child at all.

Questions for Governance Review

These questions are written for school boards, special education directors, state education agencies, vendors, professional associations and families. They are meant to be asked with a real evaluation report on the table.

FOR SCHOOL BOARDS AND SUPERINTENDENTS

1. Do we know whether the evaluation reports our teams rely on are drafted with AI, by our own staff or by contractors? Do our contracts say anything about it?
2. When a deaf or hard of hearing student is evaluated, is the evaluation administered in the child's language by someone fluent in it? If not, why not, and what would it cost to change?
3. Do our families learn, from the report itself, how it was produced?

FOR SPECIAL EDUCATION DIRECTORS

4. Does every evaluation report for a deaf or hard of hearing student include a section describing how each test was administered, through whom and with what departures, before any score is interpreted?
5. When a report recommends accommodations for a deaf or hard of hearing student, were those recommendations written by the evaluator for this child, or generated from scores?
6. For students with more than one disability, such as deafness and autism or deafness and vision loss, who on the evaluation team is qualified to separate what comes from each?
7. Do we keep the AI-generated first draft of an evaluation report with the signed report, and could we produce both if a family asked?

FOR STATE EDUCATION AGENCIES

8. Does the state have, or could it build, an evaluation center or outreach team with evaluators fluent in ASL, as Washington's Center for Deaf and Hard of Hearing Youth provides?
9. Should state guidance require that evaluation reports disclose the use of AI drafting tools and record administration conditions for every test?
10. Do state monitoring reviews check whether evaluations of deaf and hard of hearing students were administered in the child's language and mode?

FOR VENDORS

11. Does your report writer require the evaluator to record administration language, interpreter use, departures and sensory factors before it drafts interpretation? If not, would you add it?
12. When administration conditions indicate that norms do not apply, does your tool suppress standard interpretive language for those tests?
13. On how many evaluations of deaf, hard of hearing or deafblind children was your model developed or tested?

FOR PROFESSIONAL ASSOCIATIONS AND PREPARATION PROGRAMS

14. Should professional guidance treat disclosure of AI drafting in evaluation reports as required rather than a matter of judgment?

FOR FAMILIES AND ADVOCATES

15. Was my child tested in her language, by someone fluent in it, or through an interpreter? Does the report say so, and does it say what the scores mean as a result?
16. Was any part of the report drafted with software? If so, which parts, and did the evaluator write the parts that describe my child?

NOT LEGAL ADVICE

This piece describes federal special education rules, professional testing standards, published research and product claims as NCG reads them. Pine Lake, Jordan, her family and the evaluator are fictional composites, and no statement about the fictional evaluator's tool describes any named product's output. Families should contact their state's Parent Training and Information Center or an attorney, and districts their own counsel. NCG does not evaluate, endorse or rank any assessment, report writing or IEP product.

☐ Ask for the evaluation report at least a week before the eligibility or IEP meeting.

☐ Look for a section describing how each test was given: in what language, through whom, with what changes.

☐ If your child was tested through an interpreter, ask whether the report says the norms do not apply and what the scores mean instead.

☐ If your child has a vision loss as well as a hearing loss, ask whether any visual test results were affected and whether a functional vision assessment was done.

☐ Ask whether any part of the report was drafted with AI, and whether the evaluator reviewed the administration conditions before signing it.

☐ Ask whether your state has a deaf and hard of hearing evaluation center or outreach team with fluent signing evaluators.

☐ If you disagree with an evaluation that did not use your child's language, request an independent educational evaluation in writing.

TIPS

Know something about a sign language AI system that the public should know? Send it to tips@novaracg.com. Every tip is read by me and kept confidential. If you work in the UK, writing to us is not by itself a protected disclosure under the Public Interest Disclosure Act 1998, so before you share employer information, check your options with Protect (protect-advice.org.uk), the free whistleblowing advice line.

Notes

1. Heather M. Grizzle, "Six Children, One Sentence: The Special Factors a Generator Cannot Write," Novara Consulting Group, November 2026, part seven of the AI x IEP series. https://www.novaracg.com/?p=1438
2. Heather M. Grizzle, "The Screener Was Listening: AI, Child Find and the Children It Cannot Hear," Novara Consulting Group, November 2026, part eight of the AI x IEP series. https://www.novaracg.com/?p=1442
3. 34 CFR 300.303(b)(2) (reevaluation at least once every three years unless the parent and the public agency agree it is unnecessary). https://www.ecfr.gov/current/title-34/section-300.303
4. Ryan L. Farmer, Adam B. Lockwood, Randy G. Floyd and Ashley D. Sisco, "How are school psychologists using artificial intelligence in 2024? A descriptive study," School Psychology 41, no. 2 (2026): 137-149, doi:10.1037/spq0000713; summarized in American Psychological Association, "How school psychologists are using AI in practice," Article Spotlight, 2 February 2026. https://www.apa.org/pubs/highlights/spotlight/school-psychologists-ai
5. PAR, "PAR Launches AI Report Writer, Transforming the Way Psychologists Work," PAR blog, November 2025. https://www.parinc.com/learning-center/par-blog/detail/blog/2025/11/12/par-launches-ai-report-writer--transforming-the-way-psychologists-work. The fictional evaluator's tool is not modeled on any named product.
6. PAR, "AI Report Writer FAQ," read November 2026. https://www.parinc.com/ai-report-writer-faq
7. 34 CFR 300.304(c)(4) (child assessed in all areas related to the suspected disability, including, if appropriate, health, vision, hearing, social and emotional status, general intelligence, academic performance, communicative status and motor abilities) and 300.304(c)(6) (evaluation sufficiently comprehensive to identify all of the child's special education and related services needs). https://www.ecfr.gov/current/title-34/section-300.304
8. 34 CFR 300.305(a)(1) (review of existing evaluation data, including evaluations and information provided by the parents, current classroom-based, local or state assessments and classroom-based observations, and observations

by teachers and related services providers). https://www.ecfr.gov/current/title-34/section-300.305

9. American Educational Research Association, American Psychological Association and National Council on Measurement in Education, Standards for Educational and Psychological Testing (Washington, DC: AERA, 2014), Standard 6.3 (changes or disruptions to standardized administration procedures should be documented and reported to the test user).
10. 34 CFR 300.304(c)(1)(iii)-(v) (assessments used for purposes for which they are valid and reliable, administered by trained and knowledgeable personnel, in accordance with the producer's instructions) and 300.304(c)(3) (results must reflect aptitude or achievement rather than impaired sensory, manual or speaking skills). https://www.ecfr.gov/current/title-34/section-300.304
11. 34 CFR 300.306(a)(2) (copy of the evaluation report and documentation of eligibility to the parent) and 300.306(c)(1)(i) (draw upon information from a variety of sources). https://www.ecfr.gov/current/title-34/section-300.306
12. Heather M. Grizzle, "Twenty Files: Auditing the AI-Drafted IEP," Novara Consulting Group, 7 November 2026, part four of the AI x IEP series. https://www.novaracg.com/?p=1423
13. David Smith, Kristen Secora, Brittany Lee and Marie Coppola, American Journal of Speech-Language Pathology (2025), doi:10.1044/2025_AJSLP-24-00298; figures as summarized in the authors' 2026 presentation "How Prepared are SLPs to Work with Deaf Children Who Use ASL?" https://languagecreationlab.media.uconn.edu/wp-content/uploads/sites/644/2026/04/Smith_etal_2026_ACE-DHH_alignment_SLP_qualifications.pdf
14. Heather M. Grizzle, "Thirteen States Are Public. Sixty-Four Percent of Their School Units Say Nothing," Novara Consulting Group, 1 November 2026, reporting the National Deaf Education Data Index version 2.0 state files for Delaware and Hawaii. https://www.novaracg.com/2026/11/01/thirteen-states-are-public-sixty-four-percent-of-their-school-units-say-nothing/
15. Washington Center for Deaf and Hard of Hearing Youth, "Consultation and Evaluation," read November 2026. https://www.cdhy.wa.gov/outreach-services-ages-6-22/consultation-assessment-eval/
16. Visual Language and Visual Learning (VL2), Gallaudet University, "Visual Communication and Sign Language Checklist." https://vl2.gallaudet.edu/

visual-communication-and-sign-language; see also Kansas School for the Deaf, "American Sign Language (ASL) Assessments." https://ksdeaf.org/asl-assessments

17. National Center on Deaf-Blindness, state deaf-blind projects. https://www.nationaldb.org/
18. Heather M. Grizzle, "Forty-Five Days: When a Parent Asks for the AI's Draft," Novara Consulting Group, 8 November 2026, part five of the AI x IEP series. https://www.novaracg.com/?p=1429
19. Heather M. Grizzle, "Exhibit P-4: The AI's Draft Goes to a Hearing," Novara Consulting Group, November 2026, part six of the AI x IEP series. https://www.novaracg.com/?p=1434; 34 CFR 300.502(b).
20. 34 CFR 300.613(b)(1) (right to a response from the participating agency to reasonable requests for explanations and interpretations of the records). https://www.ecfr.gov/current/title-34/section-300.613
21. Heather M. Grizzle, "Drafted by AI, Signed by a Human," Novara Consulting Group, 15 October 2026, part one of the AI x IEP series. https://www.novaracg.com/2026/10/15/drafted-by-ai-signed-by-a-human/
22. Riverside Insights, "Do's, Don'ts, and Depends: Using AI Wisely in Psychoeducational Reports," webinar presented by Rebecca Branstetter, 15 September 2026. https://riversideinsights.com/training-resources/webinars/content/dos-donts-and-depends-using-ai-wisely-in-psychoeducational-reports
23. Heather M. Grizzle, "The Contract Never Asked for the Draft," Novara Consulting Group, 1 November 2026, part two of the AI x IEP series. https://www.novaracg.com/2026/11/01/the-contract-never-asked-for-the-draft/

How to cite

Grizzle, H. M. (2026, November 12). Scores Without Footnotes: AI-Written Evaluation Reports and the Deaf Child. Novara Consulting Group. https://www.novaracg.com/2026/11/12/scores-without-footnotes-ai-written-evaluation-reports-and-the-deaf-child/