

What Counts as Evidence for Gestalt Language Processing?

A Methodological Review of Measurement, Construct Validity, and Epistemic Exclusion in Autism and Language Research

Jaime Hoerricks*

PhD, Director of Research — The Praeceptory at Towcester Abbey

***Corresponding author:**

Jaime Hoerricks, PhD, Director of Research — The Praeceptory at Towcester Abbey.

ABSTRACT

People often talk about gestalt language processing (GLP) in autistic communities, therapy spaces, and classrooms, but it can be difficult to find clear recognition of it in mainstream autism research. This paper explores why that may be the case.

The paper does not argue that every idea about GLP has already been proven. Instead, it asks whether current research methods are built to recognise this kind of language development in the first place. Much autism and language research has been shaped around analytic expectations of communication. Researchers and professionals often expect language to appear quickly, in small parts, and in ways that look flexible and easy to measure from the outside.

When those expectations are treated as “normal,” autistic communication can be misunderstood. Repeated phrases may be dismissed as meaningless copying. Scripted language may be seen as immature. Delayed responses may be mistaken for lack of understanding or lack of social interest. In many cases, the behaviour itself is real, but the meaning attached to it may reflect the limits of the assessment system rather than the autistic person’s actual abilities.

This paper argues that some difficulties in studying GLP may come from the tools themselves. If assessments are designed mainly around analytic language development, they may fail to recognise other valid ways language and meaning can grow. The result is that some autistic forms of communication may become harder for institutions to see, support, and value.

Keywords: Gestalt Language Processing (GLP), Autism Spectrum Disorder (ASD), Language Development, Natural Language Acquisition, Speech-Language Pathology, Echolalia, Pragmatic Language, Evidence-Based Practice.

Received: June 23, 2026;

Accepted: July 04, 2026;

Published: July 11, 2026

Introduction

Gestalt language processing has become increasingly visible in autistic, clinical, and educational discourse, yet it remains unevenly held within the formal research literature. In some contexts, it is treated as a useful developmental framework for understanding echolalia, scripting, and language that appears to move from larger stored units toward later flexibility. In others it is regarded as underdefined, insufficiently operationalised, or not yet supported by a sufficiently mature empirical base. That tension is real. But it also

raises a more basic question than the familiar debate over whether GLP has been “proven.” When a pattern is repeatedly recognised in practice yet remains unstable in the literature, what exactly is missing—the phenomenon itself, or the methods through which the field has been trained to detect it?

This paper begins from the premise that evidence is never wholly separate from the tools that produce it. In psychological research, what becomes visible depends not only on what is present in the learner,

Citation: Jaime Hoerricks (2026) What Counts as Evidence for Gestalt Language Processing? A Methodological Review of Measurement, Construct Validity, and Epistemic Exclusion in Autism and Language Research. *J Clin Neuropsychol Prac* 1: 1-12.

but on what a task is designed to elicit, what counts as a valid response, what forms of timing and output are rewarded, and what assumptions are built into the interpretation of behaviour. Apparent inconsistency or weakness in the GLP literature cannot therefore be treated as a self-explanatory fact. In some cases, it may reflect conceptual immaturity. In others it may reflect something more methodological: a mismatch between the form of language organisation under consideration and the assessment architecture used to observe it.

GLP is useful here not only as a clinical framework but as a methodological stress test. The purpose of this paper is not to defend every claim made under the GLP umbrella, nor to treat current formulations as fully settled. The literature remains uneven, and that should be said plainly. Nor is the argument that all echolalia, scripting, or context-bound language should automatically be reinterpreted as evidence of a gestalt-organised developmental pathway. The claim is narrower and, in some ways, more demanding: current evidence hierarchies in autism and language research may be less neutral than they appear. Some of what looks like weak or inconsistent evidence for GLP may instead reflect the fact that many common measures are better suited to capturing analytic, decontextualised, and speeded forms of language performance than to detecting larger stored units, delayed segmentation, relationally anchored meaning, or coherence that emerges more clearly across time than in isolated testing moments. Under those conditions, absence in the literature may not cleanly indicate absence in the learner—it may instead indicate that the phenomenon has been placed inside an evidentiary frame that was never particularly well designed to see it.

The argument advanced here is deliberately cautious in its strongest claims. It does not argue that GLP has already been fully standardised as a singular construct across all relevant literatures, nor does it deny the reality of support needs, developmental differences, or the need for clearer operational definitions. What it does argue is that the current debate is often framed too narrowly. The question is not only whether GLP has been sufficiently validated according to existing standards. The question is whether those standards are themselves adequately designed to distinguish between lack of evidence and evidence rendered difficult to detect by construct–instrument mismatch. Where that distinction is ignored, the field risks treating one developmental architecture as though it were the neutral baseline for all others.

The deeper problem, if that is the case, is that what counts as evidence may already have been shaped by a model of language development narrow enough to make some autistic forms of coherence appear perpetually partial, weakly evidenced, or not quite real. The aim of this paper is to examine that possibility, and to argue that some of the uncertainty surrounding GLP may tell us as much about the limits of our measures as it does about the phenomenon itself.

Review Scope and Method

This paper is a narrative methodological review. Its purpose is not to calculate pooled effects, adjudicate an intervention outcome, or

claim exhaustive coverage of the autism and language literature. The question is more conceptual: how do the assumptions built into autism and language research shape what becomes legible as evidence for particular forms of language organisation? The concern is not only what the literature says about GLP, but what kinds of developmental coherence the literature is structurally equipped to recognise in the first place.

The review is accordingly organised around a problem of measurement and interpretation rather than around a single diagnostic or therapeutic framework. Method is treated here not as a neutral container for findings, but as part of the argument about how findings come to mean what they mean. Many of the behaviours relevant to GLP—echolalia, scripting, unusual phrasing, delayed response timing, literal interpretation—are not difficult to notice at the level of surface pattern. The difficulty lies in what follows: whether those patterns are treated as signs of deficit, developmental immaturity, communicative adaptation, or evidence of a different organising logic.

The paper draws on five overlapping bodies of scholarship. The first is autism diagnostic and developmental frameworks, examined for how they have historically stabilised certain surface features as diagnostically or developmentally salient. The second is literature on echolalia, scripting, and related phenomena—work that has variously framed these as nonfunctional repetition, communicative strategy, developmental resource, or evidence of a whole-to-part language profile. This literature marks one of the clearest fault lines between descriptive agreement and interpretive disagreement. The third is pragmatic and social communication assessment, with attention to how conversational timing, inferencing, literalness, and reciprocity are operationalised—domains often treated as straightforward markers of competence yet deeply shaped by assumptions about speed, sequencing, and what counts as a "social" response. The fourth is developmental language measurement more broadly, where competence is frequently inferred from structured, decontextualised, or speeded tasks. The fifth is psychometric and methodological scholarship on construct validity, ecological validity, task demands, observer-theory dependence, and the interpretive limits of behavioural proxies—sources that provide the vocabulary needed to ask not only whether a behaviour has been captured, but whether the measure has captured the construct it is presumed to represent.

Sources are included based on conceptual relevance rather than exhaustiveness, with attention to studies in which repetition, scripted language, delayed response, unusual phrasing, or literal interpretation are operationalised as markers of deficit or atypicality—especially where those interpretations proceed without sustained consideration of alternative models of language organisation.

GLP is not treated here as a singular, fully standardised construct. That lack of consensus is itself part of the problem under examination. The term is used as a practical shorthand for a family of claims in which larger stored or patterned units of language are treated as developmentally meaningful rather than as mere behavioural excess, non-generative residue, or failed approximation. The review is less concerned with defending

one model in full than with asking whether existing evidence architectures are flexible enough to distinguish between an unvalidated claim and a poorly measured one.

Across the selected literatures, the paper tracks three related questions. What kinds of behavioural performances are being treated as valid evidence of language competence? What assumptions about sequence, timing, segmentation, novelty, and context are built into those judgments? And where a behaviour associated with GLP is identified, does the literature treat it primarily as a deficit marker, or does it allow for the possibility that it may reflect a different route to coherence? These questions are not asked to reject behavioural data, but to examine the conditions under which such data are interpreted as evidence of specific underlying mechanisms.

No new data were collected and no human participants were involved; ethical approval and preregistration were not required. The contribution lies in the structured re-examination of how existing literatures define evidence and stabilise some developmental forms as more measurable than others—an argument about evidentiary fit, not a final verdict on GLP.

Behavioural Description is not Explanatory Mechanism

One of the most persistent confusions in autism and language research is the tendency to treat a stable behavioural description as though it were already a validated account of underlying organisation. Clinical and developmental frameworks appropriately begin at the level of what can be observed: repeated phrases, unusual prosody, delayed response, literal interpretation, atypical conversational timing, or language that appears overly formal, idiosyncratic, or context-bound. Such features can often be identified with reasonable consistency across observers, and in that sense, they carry real descriptive value. The problem begins when descriptive value is quietly asked to do more than it can bear—when a recurring surface pattern is treated not only as something reliably noticed, but as though it already reveals what kind of developmental or cognitive process produced it.

Reliability at the level of observation does not guarantee adequacy at the level of explanation. A clinician may accurately note that a child repeats larger chunks of language, pauses before responding, or interprets a question more literally than expected. Those observations may be entirely sound. But their meaning is not self-evident. The same repeated utterance may be read as non-generative residue, communicative strategy, affective regulation, stored language unit, or evidence of a whole-to-part developmental pathway depending on the model brought to it. Likewise, delayed response may be interpreted as impaired reciprocity, slowed processing, overload, coherence-seeking, or a timing mismatch between the learner and the demands of the interaction. In each case, the behaviour is visible. What remains contested is what kind of behaviour it is.

The difference between a description and an explanation has practical consequences. Once a behavioural pattern is taken to signify a particular internal architecture, it shapes diagnostic interpretation, referral pathways, intervention goals, and broader

judgments about competence. A child whose language is described as repetitive may be approached differently depending on whether that repetition is understood as mere symptom, partial approximation, or meaningful developmental unit. A learner whose responses arrive late may be seen as socially disengaged, cognitively delayed, or processing through a different route depending on what the observer assumes about timing and reciprocity. The surface description may be identical across these readings. The practical consequences are not.

Autism research is especially vulnerable to this slippage because so much of the field depends on behavioural inference. Diagnostic frameworks necessarily rely on observable patterns—that is not, in itself, a flaw. The difficulty arises when descriptive systems are mistaken for explanatory ones: when the fact that a pattern can be named, coded, and reliably identified is allowed to stand in for a settled account of what that pattern means developmentally. A checklist item, rating scale, or standardised observation can stabilise a phenomenon at the level of recognition while leaving its underlying organisation radically underdetermined. A learner may fail to produce isolated novel language on demand while still demonstrating patterned retention, context-sensitive deployment, or later transformation of larger stored forms. If the assessment frame is built to recognise only one route to visible competence, a descriptive shortfall can quickly become an explanatory misclassification.

The history of echolalia illustrates this clearly. For decades, echolalia was often treated as repetitive or nonfunctional language because its surface form appeared derivative rather than novel. Yet that reading was never the only one available. Prizant's work challenged the assumption that repeated language was empty residue, showing instead that echolalic speech could be communicatively purposeful, interactionally timed, and developmentally meaningful [1,2]. The observation itself—that language was repeated—did not change. What changed was the interpretive frame. The same surface feature, once understood differently, no longer functioned as straightforward evidence of communicative failure. This is precisely the kind of shift that reveals how fragile the boundary can be between reliable description and overconfident explanation.

The same logic applies across the features considered in this paper. Scripted or formal language, literal interpretation, and apparent reciprocity difficulties all depend on prior decisions about what counts as appropriately flexible, adequately timed, or recognisably responsive. In each case, the field may be seeing something real while misidentifying the level at which that realness resides. A behavioural indicator is a pattern that can be observed and described. A developmental mechanism is an explanatory claim about how that pattern emerges and what trajectory it may represent. When those two levels are collapsed, behaviours that are straightforward to observe become proxies for architectures that have not been directly measured—and the resulting inferences take on the appearance of empirical inevitability simply because the surface pattern itself is real. Where the dominant tools of observation are additionally shaped by one developmental logic more than another, that problem deepens. That is the issue to which the next section turns.

Construct Validity and the Hidden Analytic Norm in Language Assessment

The methodological problem at the centre of this paper is not simply that some autistic language profiles remain under-described. It is that many of the tools used to describe them may be narrower than they appear. In psychological and developmental assessment, a measure is often treated as valid when it reliably captures the construct it claims to assess. Yet construct validity is not established by consistency alone. A task may be highly standardised, highly replicable, and even clinically useful while still representing only a partial or distorted slice of the phenomenon it purports to measure. This distinction becomes especially important in language assessment, where what counts as "competence" is often shaped as much by task design as by the learner's actual organisation of meaning.

The clearest form of this problem is construct underrepresentation. A language task may successfully capture the ability to produce segmented verbal output under externally structured conditions while failing to capture other forms of organisation that are no less developmentally meaningful. A learner who stores and deploys larger patterned units, depends more heavily on contextual anchoring, or shows coherence more clearly across time than in isolated elicitation may appear weaker on a measure that is not actually broad enough to represent the full range of what "language competence" can be. The score may be accurate in a narrow sense while still misrepresenting the learner's underlying developmental profile.

At the same time, some measures risk construct contamination: they do not merely sample language, but load heavily on auxiliary demands that are then mistaken for language itself. A task may appear to assess pragmatic skill while also requiring rapid processing under social pressure. It may appear to assess inferencing while depending on familiarity with culturally patterned figurative language. It may appear to assess expressive competence while presuming tolerance for decontextualised prompts, comfort with externally imposed timing, or readiness to generate novel output in a form that fits the examiner's expectations. When those extra demands are ignored, the resulting interpretation can overstate deficit in the named domain.

This matters because autism research has often treated decontextualisation as a route to precision. Standardisation aims to reduce noise—to strip away contextual variation so that the same construct can be measured across learners. That logic has value. But in language assessment, context is not always noise. It may be part of the phenomenon. If a learner's language is strongly shaped by relational safety, sensory regulation, familiar patterns, or stored units whose meaning depends on prior interactional history, then removing those conditions does not necessarily purify the construct. It may instead deform it. What looks like a cleaner measure may be a narrower one, and what appears as poor performance may sometimes reflect the disappearance of the very supports through which the learner's competence becomes visible.

This is where the hidden analytic norm becomes most consequential. Many dominant measures of language

development are built around an implicit developmental logic in which smaller units precede larger ones, segmentation precedes flexibility, novelty is privileged over patterned reuse, and rapid adaptation under externally structured conditions is treated as a hallmark of maturity. That sequence is so familiar in mainstream developmental discourse that it often passes as common sense rather than theory. Yet it is not theory-free. It is a model of language organisation—one that may describe many learners well, but which is still a model rather than a universal law. When assessment tools are calibrated to that model, they are not simply measuring language in the abstract. They are measuring how readily a learner's language becomes legible within an analytic frame.

The consequence is not that such tools become useless, but that their apparent neutrality becomes harder to defend. Over time, reliance on tasks that privilege segmented output, rapid novelty, and decontextualised performance means that learners whose competence appears most clearly in other forms will be predictably underrepresented in the data. That underrepresentation can begin to look like empirical confirmation of the model itself—analytic forms of language appearing robustly measurable because the tools are designed to detect them, while other forms appear inconsistent or weakly evidenced because they are harder to stabilise within the same evidentiary architecture. At that point, the field is no longer simply measuring heterogeneity. It is participating in the production of it.

Empirical caution cuts in more than one direction. It is not only a caution against overclaiming new developmental frameworks. It is also a caution against assuming that familiar measures already capture the construct in full. Where language assessment is concerned, the most consequential methodological error may not always be false inclusion. It may also be false exclusion—the quiet exclusion of forms of competence that do not appear in the shape the instrument was built to recognise. The question is therefore not whether analytic measures ever detect real difficulty—of course they do—but whether the field has been sufficiently explicit about the developmental logic those measures embed. If it has not, then what presents as neutral assessment may be a narrower epistemic filter than it appears. The next step is to return to the specific behaviours that keep being coded—echolalia, scripted phrasing, delayed response, literalness—and ask whether the dominant interpretations attached to them are genuinely warranted, or merely familiar.

Re-reading Commonly Coded Indicators through a GLP Lens

The methodological concerns outlined above become clearest when the literature returns to the same small cluster of observable features and treats them as though their meaning were already settled. Certain patterns recur with striking regularity in autism and language research: repeated phrases, scripting, unusually formal or idiosyncratic language, delayed response timing, and literal interpretation. These are often accurately noticed. The question is not whether such patterns exist, but whether the interpretive frameworks attached to them are sufficiently broad to distinguish deficit from difference, or developmental difficulty from a mismatch between the learner and the demands

of the task. The purpose of this section is not to relabel every atypical language feature as evidence of GLP, but to ask whether the dominant readings attached to several commonly coded indicators remain warranted once the possibility of gestalt-organised language is taken seriously.

a. Echolalia and Scripting: Repetition or Stored Language Units?

Few features in autism language research have been more consistently observed—or more consistently overinterpreted—than repetition. Echolalia has long occupied a peculiar place in the literature: immediately noticeable, often clinically salient, and historically burdened by an interpretive shortcut in which repeated language is treated as self-evidently less meaningful than apparently novel production. That shortcut has proven durable in part because repetition is easy to recognise at the level of surface form. But the interpretive leap from repeated form to non-generative language has never been as secure as its familiarity suggests.

Repetition, on its own, is not an explanation. It is a description of form. To say that language is repeated tells us something real about its visible structure, but not yet what developmental work that language is doing—whether it is functioning as residue, regulation, communication, rehearsal, or transformation. For many years, repeated language in autism was treated as stereotyped or nonfunctional speech, implicitly contrasted with the more highly valued marker of spontaneous, segmented, novel production. Yet that contrast rested less on direct evidence about underlying organisation than on a prior assumption that true language competence ought to appear as immediate, flexible novelty.

Importantly, the idea that larger stored units may play a meaningful role in language development did not originate in autism scholarship. Peters argued that early language learning may sometimes proceed through larger, initially unanalysed chunks rather than exclusively through the progressive assembly of smaller units—challenging the assumption that segmented, part-to-whole acquisition is the sole or default developmental pathway [3]. Even where the eventual endpoint includes analysis and recombination, the route may begin with wholes that are carried before they are fully parsed. The presence of formulaic or repeated language is not, in itself, evidence against development. It may indicate that development is occurring through units larger than the observer expects.

Within autism research, Prizant's work marked a crucial turning point. He argued that echolalia should not be reduced to empty repetition, but understood in relation to communicative function, interactional timing, and developmental process [1,2]. What appeared on the surface as repetition could serve multiple purposes: turn-taking, self-regulation, affirmation, request, protest, or participation in an interaction whose conventional cues were being navigated differently. The observation itself did not change. What changed was the refusal to assume that derivative form necessarily implied derivative meaning. Once the possibility of larger stored units is admitted—as Peters established in general language development and Prizant

extended within autism—repetition can no longer function as straightforward proof of communicative absence. It becomes a site where the observer must ask what kind of organisation the repeated material may represent.

Scripting presents a related but distinct problem. In clinical and educational contexts, "scripting" typically refers to familiar lines from media, rehearsed phrases, or patterned verbal sequences that appear contextually unusual or only partially adapted to the present moment. These descriptions are not necessarily wrong, but they often stop at the level of recognisable form. A phrase may be labelled a script because it is borrowed or repeated, yet that label does not tell us whether the utterance is being used rigidly, relationally, strategically, or as a transitional form in a larger process of segmentation and recombination. The term can become a descriptive convenience that smuggles in an explanatory assumption: that because the language is recognisably sourced elsewhere, it is therefore less authored or less developmentally significant.

From a GLP perspective, that assumption is difficult to sustain. If larger units of language may be stored and carried as meaningful wholes, a "script" is not necessarily the opposite of generative language. It may be a developmentally available unit whose internal segmentation is partial or emerging, but whose use is nevertheless communicative and organised. The key question is not whether the utterance originated elsewhere, but what the speaker is doing with it now—whether it shows mitigations, substitutions, prosodic shifts, or gradual reorganisation over time. These are empirical questions, but they cannot be asked if the presence of repeated or borrowed form has already been taken as evidence that the utterance is developmentally lesser.

The point is not to flatten all repetition into competence any more than earlier frameworks flattened it into deficit. Repetition is an insufficiently specific behavioural description to bear the explanatory weight it has often been asked to carry. Once that is acknowledged, echolalia and scripting cease to function as self-interpreting symptoms and become sites of methodological humility—places where the literature must ask whether it has been observing the form accurately while misunderstanding the unit of organisation.

b. Idiosyncratic or Formal Speech: Pragmatic Error or Patterned Precision?

If echolalia and scripting are often misread because the language is recognisably sourced, idiosyncratic or overly formal speech presents a related but distinct problem: the language may be clearly generated by the speaker, yet still treated as suspect because it does not sound like the kind of language the interaction expected. Autism literature has long noted patterns variously described as pedantic, formal, unusual, bookish, overly precise, or neologistic. Language that appears grammatically complex or lexically rich may nevertheless be marked as pragmatically odd because it is judged to be out of register, overly elaborate, or socially misaligned. As with repetition, the observation may be real. The question is what explanatory burden it can legitimately bear.

Historically, this phenomenon has been framed through the language of abnormality rather than architecture. Volden and Lord treated neologisms and idiosyncratic language as notable features in autistic speakers, while also finding that such productions increased alongside greater language complexity—a finding that already complicates any simple deficit account [4]. Unconventionality did not merely appear where language broke down; it could also appear where language was becoming more elaborated. Ghaziuddin and Gerstein's work on "pedantic" speaking style helped stabilise the construct in clinical discourse, but also illustrates the limits of the older frame: once "pedantic" becomes a clinically recognisable style, its recognisability can be mistaken for explanation [5]. Yet describing speech as overly formal or unusual in tone does not tell us whether the speaker is evidencing impaired social understanding, a different weighting of communicative priorities, reliance on stored patterned language, or a more effortful route to discourse planning under interactional load.

Luyster, Zane, and Wisman Weil offer a more useful distinction, separating non-generative unconventional language (such as immediate or delayed repetition) from generative unconventional language, including neologisms, idiosyncratic phrases, and pedantic speech [6]. That distinction matters considerably here. Some of the speech historically coded as odd or pragmatically weak is not derivative residue or failed approximation—it may be generated from the speaker's own linguistic system, even where its surface form departs from conventional expectations. The issue is no longer whether the language is "normal," but what kind of organisation it reflects and how the interaction is rewarding or penalising its visible form.

The central methodological problem is that pragmatic mismatch and patterned precision can look similar from the outside. A speaker may choose unusually formal vocabulary, provide more detail than the interaction seems to require, or produce a phrase that feels semantically exact but socially unexpected. Under a narrow pragmatic frame, such speech is easily read as failure of audience design or register selection—and sometimes that reading may be warranted. But it is not the only plausible reading. The speaker may be optimising for semantic specificity rather than social economy, or what appears as "too much" language may function as compensatory explicitness where prosodic or contextual cues are less readily available and lexical precision is doing more of the interactional work.

Several strands of the literature support this more heterogeneous reading. Diehl et al. found that autistic adolescents were less likely to use prosodic cues to resolve syntactic ambiguity—suggesting that unusually explicit or lexically over-specified language may sometimes represent adaptation rather than mere oddity [7]. Whyte and Nelson showed that pragmatic and nonliteral language development in autistic children follows different trajectories, cautioning against any assumption that surface unconventionality transparently reflects a single deficit [8]. Reindal et al. similarly emphasised the interplay between structural and pragmatic language differences in children evaluated for autism, underscoring that pragmatic interpretation cannot be cleanly separated from broader language architecture

[9]. More recently, Wu, Lam, and To (2025) found that Chinese autistic adults produced more neologisms, idiosyncratic phrases, and pedantic language than non-autistic peers in spontaneous speech, but the pattern did not reduce to simple grammatical weakness—it appeared to involve a more complex interaction between generative language, fluency, and discourse-level processing demands [10]. Zane and Luyster have argued that neologisms and related productions may point toward an autistic "linguatype," suggesting these forms are not merely errors to be subtracted from competence but potentially meaningful clues about how competence is being organised [11]. Trayvick et al. note that the field still relies on qualitative coding and inconsistent definitions in this area—a reminder that if the category itself is loosely defined, the confidence with which it supports developmental inference should be correspondingly restrained [12].

The most defensible conclusion is neither that formal or idiosyncratic language should be pathologised by default, nor that it should be automatically reclassified as hidden sophistication. These speech patterns are heterogeneous, and the dominant literature has often been too quick to treat them as transparently indexing pragmatic weakness. What is being coded may include several overlapping phenomena—register mismatch, semantic overprecision, partial scripting, compensatory explicitness, or genuinely unconventional lexical generation. Those are not interchangeable. From a GLP perspective, unusual register may not be peripheral to the architecture but one of the places where the architecture becomes visible. A phrase that sounds too formal for the moment may reflect not absent social awareness but the fact that the speaker's most available, most precise linguistic unit is not the one the interaction expected. The issue is not whether the speech is unusual. It is whether unusuality has been made to stand in for inferential certainty.

c. Conversational Latency: Delayed Response or Delayed Legibility?

Conversational delay is among the most routinely described yet thinly theorised features in autism language research. Clinical observations and assessment protocols frequently record that an autistic speaker pauses before responding, or requires additional time before producing language that fits the immediate conversational turn. Such latency is often coded as reduced reciprocity, slowed processing, or difficulty maintaining conversational flow. The problem arises when the presence of delay is treated as though it transparently reveals a single underlying mechanism.

Latency is first and foremost a temporal description. It tells us when language becomes externally visible relative to the demands of the interaction—not what internal work the pause represents. A delayed response may reflect difficulty formulating language, but it may also reflect the time required to locate, adapt, or partially segment a larger stored unit before producing it in a form that fits the present exchange. From the outside, both processes produce the same visible pattern: silence followed by speech. The difference lies not in the timing itself, but in the organisation of the language that precedes it.

Hoericks argues that many interpretations of conversational delay involve a category error: the assumption that response latency necessarily reflects developmental lag rather than differences in the architecture of language organisation [13]. Under analytic models, competence is measured by rapid retrieval and recombination of smaller linguistic units, so delayed output appears naturally as slowed or inefficient processing. But if language is sometimes organised in larger patterned units whose segmentation or adaptation occurs closer to the moment of use, then additional time before speaking may represent a different internal process altogether—not the absence of language, but the work of transforming stored coherence into locally appropriate form.

This distinction matters because conversational timing is rarely assessed in isolation. Measures of pragmatic competence frequently assume that timely response is itself evidence of social understanding, such that a speaker who pauses risks being interpreted as disengaged or uncertain. Yet the social meaning of timing depends on the assumptions built into the interaction. If the conversational frame presumes rapid analytic retrieval, delayed responses will appear deviant even where the eventual utterance demonstrates coherence, relevance, and precision. The judgment may reveal as much about the temporal expectations of the assessment frame as about the developmental status of the speaker. Work on prosody and ambiguity resolution, pragmatic development, and structural-pragmatic interactions in autism all caution against treating conversational timing as a straightforward proxy for communicative competence [7-9].

Within a GLP framework, latency may signal the moment at which stored language meets the requirements of the present interaction. A speaker drawing on larger patterned units may need additional time to identify a fitting unit, adjust it through partial segmentation or substitution, or reconcile the relational meaning carried by a stored phrase with the demands of the immediate exchange. That work appears from the outside only as a pause—but the pause may be doing real developmental labour. Where analytic production emphasises speed of retrieval, gestalt-oriented production may emphasise coherence before articulation, a difference the interaction registers as delay even when the resulting language is communicatively effective.

None of this eliminates the reality that conversational timing can present genuine challenges. Rapid turn-taking environments, classrooms, and formal assessments reward immediate response in ways that leave little space for slower language production, and under those conditions latency can function as a genuine barrier to participation. But recognising the barrier is not the same as assuming it reveals a deficit in language competence. A pause is an observable fact. What it means depends on the theory brought to it—and if the possibility of larger language units and delayed segmentation remains open, the same pause becomes interpretable in multiple ways. Conversational delay is a point where the architecture of language and the tempo of the interaction meet, and sometimes misalign. That misalignment is easy to misread precisely because the language that eventually arrives may look entirely ordinary. The only unusual feature is the time it took to appear.

d. Literal Interpretation: Inferential Deficit or Linguistic Commitment?

Literal interpretation is one of the most familiar and least carefully theorised features in autism language research. It is frequently described as evidence of pragmatic weakness or reduced flexibility in inferencing—autistic speakers "taking things literally," particularly with idioms, sarcasm, figurative language, or indirect requests. At the level of surface description, this observation is often accurate enough. The problem, as with the features considered above, is that the behavioural description is treated as though it already settles the question of what kind of process is being observed.

"Literalness" is often assumed to transparently indicate a deficit in higher-order social understanding. Yet many tasks used to assess nonliteral or pragmatic language do not isolate a single inferential construct. They bundle multiple demands at once: rapid contextual integration, familiarity with culturally patterned idioms, sensitivity to prosodic or facial cues, tolerance for ambiguity, and willingness to override surface semantics in favour of an implied meaning that is context-dependent and socially compressed. When a speaker fails such a task, the result may reflect difficulty with nonliteral interpretation—but it may also reflect differences in cue weighting, processing tempo, cultural exposure, or task design. Whyte and Nelson showed that pragmatic and nonliteral language in autistic children follow distinct developmental trajectories, reinforcing that nonliteral comprehension is not a binary skill present or absent at a single threshold [8]. Reindal et al. similarly underscore the entanglement of structural and pragmatic language differences, complicating any attempt to treat figurative performance as a pure index of social cognition [9]. Diehl et al. further illustrate the point: where meaning depends on the integration of multiple cue streams, reduced use of one stream can alter the apparent outcome without indicating absence of interpretive capacity in the domain itself [7].

From a GLP perspective, literal interpretation may be especially vulnerable to misreading because figurative language often depends on forms whose meaning is distributed across prior exposure, relational context, and patterned usage rather than transparent semantic composition. An idiom is not simply a sentence plus an inference. It is a stored social-linguistic object whose meaning must be learned as a unit, recognised in context, and weighed against its surface lexical content. For a speaker whose language processing depends heavily on larger stored units or strong commitment to the integrity of lexical meaning, the demand to abandon surface semantics quickly in favour of a compressed social reading may not represent a failure of reasoning so much as a different route to meaning—less an inability to infer than a different threshold for when inference is warranted.

This is where the distinction between inferential deficit and linguistic commitment becomes useful. A speaker may appear literal not because they cannot imagine alternative meanings, but because the utterance has not yet met the conditions under which abandoning its semantic surface feels justified. In some cases, this reflects limited familiarity with the patterned form of an idiom.

In others it reflects a principled weighting of lexical precision over social convenience—particularly where implied meanings have proven unstable or socially costly. A stronger commitment to explicit semantic content may function not as impoverishment but as a stabilising strategy. The apparently "literal" reading is not necessarily a failure to go beyond the words. It may be an insistence that the words themselves remain accountable.

Many pragmatic tasks assume that mature language use involves efficient departure from surface meaning when context invites it. But context in these tasks is often artificially compressed, the inferential pathway treated as obvious because it is obvious to the test designer. A speaker who hesitates, asks for clarification, or holds the lexical meaning longer than expected can be scored as deficient even where the response reflects careful semantic monitoring rather than absent comprehension. As Luyster et al. caution, many spoken language features in autism have been historically grouped in ways that obscure whether they are generative, context-sensitive, or developmentally meaningful—and literal interpretation is no exception [6]. What is coded under that single label may include unfamiliarity with a particular idiom, delayed access to nonliteral mappings, context-specific caution, preference for explicitness, or genuine difficulty with pragmatic inference. These are not interchangeable, yet once the surface response is coded as "literal," the literature can behave as though the explanatory work is already done.

A literal response can be described. What it means developmentally remains an interpretive question—and one that the field has answered too quickly for too long.

Methodological Implications: What Would Count as Evidence?

If the problem is not simply that GLP remains contested but that many dominant tools of autism and language research are calibrated to one developmental logic more than another, then the solution cannot be reduced to calls for "more evidence" in the abstract. More studies conducted within the same evidentiary architecture may only reproduce the same interpretive narrowing with greater statistical confidence. The question is what kinds of research design would be capable of detecting gestalt-organised language without forcing it to appear in analytically preferred form first—a question that requires attention to ecological validity, temporal scale, task demands, coding practices, and inferential discipline.

a. Ecological Validity and Naturalistic Sampling

Evidence for gestalt-organised language is unlikely to emerge reliably from designs that remove the very conditions under which such language becomes visible. A speaker whose language depends on familiar routines, emotionally salient scripts, co-regulated interaction, or sufficient time to transform larger stored units may appear markedly less capable in a stripped-down elicitation task than in ordinary conversation or repeated interaction with a familiar partner. If the goal is to determine what language architecture is present—rather than only what performance can be elicited under analytic pressure—then the context of sampling becomes part of the construct rather than an inconvenient variable to be eliminated.

This is not an argument against standardisation, nor a call to replace structured assessment with anecdote. It is a call to be more precise about what a given task can legitimately claim to measure. A decontextualised naming or inferencing task may tell us something useful about performance under those specific conditions. It may not tell us how language is organised across time, across relationships, or across contexts in which larger patterned units become available. The field has often treated structured elicitation as a neutral window into competence rather than as a specific interactional arrangement that privileges some forms of output over others. In the context of GLP, this creates a predictable distortion: the language easiest to capture in controlled tasks is the language already closest to analytic expectations.

More adequate evidence will therefore require naturalistic sampling designs that are systematic rather than merely impressionistic—repeated audio or video samples from ordinary conversation, play, collaborative activity, or family interaction, in which language can be observed across varying levels of demand and familiarity. The point is not simply to collect "real life" language but to preserve access to the contextual cues that may shape how stored units are activated, how scripts are adapted, and how unconventional or delayed forms of competence emerge when the interaction is not forcing immediate atomised output. Such sampling can be rigorous: transcribed, coded, compared across contexts, and analysed longitudinally. What changes is not the presence of method but the willingness to treat context as constitutive rather than contaminating.

Trayvick et al. note that many observed speech and language features in autism have historically depended on qualitative description and inconsistent coding, and their call for more systematic analysis of spontaneous speech—including through natural language processing—points in a promising direction [12]. The caution, however, is important: new tools do not become neutral merely by becoming computational. If NLP pipelines are trained on constructs already shaped by analytic assumptions, they may scale the same interpretive biases more efficiently rather than correcting them. The same applies to observational coding more broadly. As Luyster et al. demonstrate, much turns on whether spoken language features are classified as non-generative, generative, or functionally meaningful [6]. A naturalistic sample is not automatically more valid if the coding frame remains too coarse. Ecological validity is not only about where language is collected—it is also about whether the coding system can preserve distinctions that matter developmentally.

b. Longitudinal and Transformation-Sensitive Designs

Cross-sectional snapshots are poorly suited to detecting the kind of developmental signal that GLP involves. A single assessment session or tightly bounded experimental task may capture whether a particular form of language is visible at a given moment, but not how language reorganises across time—which is precisely where the evidence may lie. Many features associated with GLP are dynamic rather than static: a phrase that appears as repetition at one moment may later show mitigation, substitution, or contextual narrowing; a stored script may gradually yield smaller units available for recombination;

a response that initially arrives late may show greater fluency under familiar conditions while remaining effortful under decontextualised demand. Research designs built to detect only whether a behaviour is present or absent at a single time point risk missing the developmental movement that would distinguish a fixed deficit from an evolving language system.

More adequate evidence will therefore require longitudinal designs that are explicitly sensitive to transformation rather than merely accumulation. Traditional developmental studies track change in terms of vocabulary size, mean length of utterance, or frequency reduction in particular behaviours. Those metrics remain useful but are insufficient if the key developmental phenomenon involves reorganisation of linguistic units rather than simple increase or decrease in output. A transformation-sensitive design would ask not only whether repeated language becomes less frequent, but whether repeated forms become more variable, more contextually specific, more internally modified, or more clearly linked to emergent segmentation—treating development as a shift in form-function relation, not merely a change in frequency counts.

The practical consequence is significant. A child who continues to use repeated or patterned language across multiple time points may appear, under a flat coding system, to be not progressing. Yet if those repetitions are becoming shorter, more selectively deployed, or increasingly recombined, the developmental picture is entirely different. Likewise, latency treated as a stable deficit marker may, if examined across contexts and partners, reveal systematic variation that signals adaptation rather than stasis. The problem is not only that cross-sectional designs miss growth—it is that poorly specified longitudinal designs can miss it too if they track only the persistence of a visible feature without examining what is changing within it. As Hoerricks argues in relation to conversational delay, what may be misread as developmental persistence can in fact be the visible trace of ongoing reorganisation [13]. Duration without sensitivity is not enough.

In practical terms, future research should combine repeated sampling with coding schemes capable of detecting mitigation, recombination, contextual redistribution, partner effects, and timing shifts under varying demand. Case-series and small-cohort designs are likely to be especially valuable in the near term—not because large samples are unimportant, but because early-stage construct clarification requires finer-grained temporal observation than broad outcome studies can easily sustain. Once the field knows what transformations to look for, larger-scale designs can test those patterns more systematically. Scaling up before the construct is adequately specified risks producing more data without producing more understanding.

c. Task Redesign: Timing, Context, Scaffolding, and Multi-Pass Elicitation

Task design determines what kinds of language are made available for observation in the first place. In autism and language research, the structure of a task often functions as a hidden theory of competence: a prompt requiring immediate response assumes rapid retrieval is part of the construct; a decontextualised question

assumes meaning can be accessed independently of relational anchoring; a single elicitation assumes first-pass output is the best index of what the speaker knows. Where the aim is to detect language organisation that may depend on larger stored units, delayed segmentation, or contextual support, task design must become more explicitly responsive to those possibilities.

The first implication concerns timing. Many conventional assessments implicitly penalise delay by treating a quick response as more competent than one that arrives after a pause, even where time limits are not formally scored. More adequate task design would distinguish between response availability and response quality—documenting whether scoring depends on first-pass immediacy, allowing longer response windows where theoretically justified, and analysing whether delayed responses differ qualitatively from immediate ones rather than collapsing both into a single performance metric.

The second implication concerns context anchoring. Stripping away contextual supports standardises the task while potentially making the construct less representative of how language is actually organised. If a speaker's access to meaning depends on familiar routines, emotionally salient themes, or concrete situational framing, then context-free prompts may measure tolerance for abstraction as much as language competence. Research designs should therefore explicitly compare performance across levels of contextual support—contrasting responses to abstract prompts with responses to anchored prompts that include visual scenes, familiar routines, or collaboratively established topics—to determine whether a behaviour reflects a stable limitation or a context-sensitive access problem.

The third implication is scaffolding, understood here not as therapeutic prompting but as a methodological variable. Conventional assessments treat scaffolding as contamination because it reduces independence. But from a developmental perspective it can be diagnostically informative. If a brief rephrase, a semantic cue, or a shift to co-constructed dialogue substantially changes the quality of a response, the task has revealed something important: not simply whether the speaker can produce a target form unassisted, but what conditions make that form available. The goal is not to inflate scores but to map the relationship between task conditions and response availability more precisely than a binary correct/incorrect outcome allows.

This leads to the strongest design recommendation: multi-pass elicitation. Many assessments rely on a single response opportunity, treating the first answer—or absence of one—as the definitive sample of competence. If some language systems require more time, more context, or more than one route into the same content, single-pass tasks are structurally biased toward analytic immediacy. A multi-pass design preserves the initial response as meaningful data while introducing one or more systematically varied follow-up opportunities—the same item presented first in standard form, then revisited with additional time, contextual framing, or minimal scaffolding. Such designs can remain rigorous: passes can be standardised, pre-specified, and theoretically motivated, with researchers defining in advance what scaffolds are permitted, what counts

as a transformed response, and how later responses are scored. A participant who produces no immediate response but offers a contextually precise, internally organised response on a second pass is not simply "better with help"—they may be demonstrating a language system whose access conditions differ from the assumptions built into the first task. Without the second pass, that distinction remains invisible.

A further consideration concerns partner effects. Language performance often varies significantly depending on whether the interlocutor is familiar, relationally safe, or less demanding in pacing. Standardisation treats partner variability as noise, but for some constructs it may be part of the signal. Comparing performance with a familiar partner and an unfamiliar examiner can reveal whether apparent difficulty is stable across contexts or amplified by the social form of the assessment itself.

d. Coding Reform: Distinguishing Form, Function, and Developmental Inference

Even where researchers improve ecological validity, extend temporal observation, and redesign tasks, a central problem remains: the meaning of a language sample still depends on how it is coded. Many of the most persistent interpretive errors in autism and language research arise not at the point of observation but at the point where behavioural description is converted into analytic category. A repeated phrase becomes "echolalia," then "non-generative language," then evidence of limited flexibility. A delayed response becomes "latency," then "pragmatic difficulty," then evidence of impaired reciprocity. An unusual utterance becomes "pedantic," then evidence of pragmatic deviation, then a proxy for broader communicative impairment. The shift from description to explanation can happen so quickly that it ceases to feel like an inference at all—and once these layers are fused, interpretation becomes circular: the behaviour is taken as evidence for the very construct that defined it.

A more defensible approach would separate coding into at least three analytically distinct layers: form, function, and developmental inference. Form refers to what can be directly observed in the language sample—immediate repetition, delayed reuse of a prior phrase, unusual lexical choice, high-register phrasing, delayed response onset, or literal uptake of a nonliteral prompt. These descriptions should be as behaviourally precise and theory-light as possible. Function refers to what the utterance appears to be doing in the local interaction: maintaining turn-taking, regulating uncertainty, preserving precision, buying processing time, signalling affiliation, indexing a stored script, or resisting an underspecified demand. Functional coding necessarily remains more interpretive, but it should be grounded in context and kept distinct from developmental mechanism. Developmental inference is the strongest and most speculative layer: the claim that the observed pattern reflects a particular kind of underlying organisation—limited generativity, pragmatic impairment, atypical prosodic integration, or possible reliance on larger stored linguistic units.

This three-layer distinction matters because the same surface form can support multiple plausible functions, and the same function can arise from different underlying mechanisms. A

repeated phrase may function as non-generative repetition in one context, and in another as a highly organised entry point into a longer response, a means of preserving precision under load, or the visible remnant of a larger stored form undergoing gradual transformation. A literal response may reflect difficulty with inferencing in one instance and in another a linguistically disciplined commitment to the most semantically stable reading when context is underspecified. Once form, function, and inference are collapsed into a single category, these distinctions become difficult to recover.

Several strands of the literature already point toward the need for greater coding precision. Luyster et al. explicitly resist overly blunt classification of spoken language features, while work on neologisms and unconventional language illustrates that not all atypical forms are equivalent in structure, function, or developmental significance [4,6,10,11]. Even studies retaining more traditional terminology reveal how much interpretive weight broad labels carry [5,9]. The lesson is not simply that terminology should be modernised, but that codebooks should be designed so that descriptive categories do not silently predetermine explanatory conclusions.

In practical terms, future studies should report coding schemas in a way that makes these layers explicit. Rather than treating "echolalia" as a terminal category, a study might first code whether the utterance is immediate or delayed, exact or mitigated, tied to prior discourse, and whether its use shifts across contexts or time points. A second layer could address probable interactional function, with multiple raters and contextual notes. Only after those steps should developmental interpretations be advanced—and even then, as hypotheses rather than as properties already contained within the behaviour. The same logic applies to "pedantic," "idiosyncratic," "literal," or "pragmatically odd." The more heavily a term has historically functioned as both descriptor and judgement, the more carefully it should be unpacked.

This is not a call for infinite relativism or suspended interpretation. Research requires categorisation, and developmental science necessarily makes inferences beyond what is directly visible. The point is that inference becomes more trustworthy when the steps leading to it remain inspectable. When codebooks preserve the distinction between what was observed, what it appeared to do, and what mechanism is being proposed, disagreement becomes more scientifically productive. In a field where constructs are contested and developmental pathways may be multiply organised, that transparency is part of what makes a claim testable.

Toward a GLP-Sensitive Research Programme

The immediate task is not to replace one orthodoxy with another, nor to insist that every instance of echolalia, scripting, delayed response, or unconventional language should be retroactively reclassified as evidence of GLP. The more disciplined position is narrower: the field's existing evidentiary routines have not yet been well designed to determine the question cleanly, and that limitation has consequences both for research and for the interpretive habits that research authorises downstream. What

follows outlines five priorities for a research programme capable of addressing that limitation.

The first is construct clarification. GLP is currently discussed across overlapping but not fully aligned domains—clinical observation, educational practice, parent report, autistic self-description, and intervention-oriented developmental models. The construct may be experientially recognisable and practically useful in some settings while remaining under-specified by the standards of mainstream psychological research. Rather than treating that mismatch as grounds for dismissal, it should be treated as a familiar early-stage scientific problem. The work required is not the production of a single reductive checklist, but clarification of which recurrent patterns—delayed but coherent response, mitigated repetition, context-bound script use, gradual recombination, transformation of larger stored forms—are most theoretically central, which are secondary, and which may be epiphenomenal or non-specific. Small-cohort and case-series designs are likely to be especially important here. When the construct itself remains partially unresolved, large samples can create an illusion of robustness while merely scaling imprecise categories. Richly sampled, longitudinal, multi-context designs that preserve the grain of the phenomenon long enough for the construct to sharpen are a more productive starting point. Scaling up before the construct is adequately specified risks producing more data without producing more understanding.

The second priority is the development of shared observational and coding frameworks that can travel across disciplines without collapsing into the bluntness that has historically characterised much autism language research. Observations that are highly meaningful in classrooms, therapy rooms, or family interaction are often translated into research categories that flatten what made them significant in the first place. The field will need coding systems that can distinguish repeated form from interactional function, surface novelty from patterned precision, and persistence from transformation. This is not simply a technical challenge—it is a translational one, requiring methodological bridges between practitioners who notice patterns in lived interaction and researchers trained to distrust anything that arrives without prior operationalisation. If those worlds remain split, the research literature will continue to privilege what is easiest to count over what may be most developmentally revealing.

The third priority is cross-linguistic and cross-cultural extension. Much of the existing literature on autism and language remains shaped by English-dominant assumptions about pragmatics, lexical flexibility, figurative language, and what counts as communicative appropriateness. If GLP involves patterns of storage, segmentation, repetition, mitigation, or context dependence, those patterns may present differently across languages with different morphosyntactic structures, discourse conventions, prosodic systems, or expectations around indirectness and literalness. Wu et al. already suggest that unconventional language phenomena in autism cannot be assumed to behave identically across linguistic contexts [10]. A GLP-sensitive research programme should resist treating English-speaking developmental expectations as universal

baselines—a concern that is especially pertinent for a journal whose scope explicitly invites attention to psychological phenomena across diverse linguistic and cultural settings. Construct clarification must begin with cross-linguistic humility rather than arrive there later as an afterthought.

The fourth priority is methodological pluralism with inferential discipline. The future of GLP research is unlikely to be served by a single privileged design. Some questions will require structured tasks, others naturalistic corpora, others longitudinal observational work or computational modelling of spontaneous language. The danger lies not in plurality but in slippage—treating methods as interchangeable when they are sensitive to different layers of the construct. A rigorous programme would be plural in design but explicit about inferential reach, stating what each method can and cannot show rather than allowing narrow tasks to make broad developmental claims.

Finally, such a programme must take seriously the fact that research in this domain does not remain confined to journals. The interpretive habits established in the literature travel outward, shaping diagnostic reasoning, assessment practices, educational expectations, and families' understanding of what kind of language they are seeing. When a surface feature is too quickly treated as proof of deficit, the cost is not merely conceptual—it becomes institutional. Children are taught through models that may misdescribe the architecture of their language. Adults are read through deficit categories that may have followed them for decades. Researchers inherit coding traditions that feel objective because they are old, not because they have been sufficiently tested against alternative developmental possibilities. A GLP-sensitive research programme is therefore not only a matter of better construct validity. It is also a matter of reducing the downstream harms that follow when narrow measurement regimes are mistaken for neutral descriptions of human development.

The most scientifically responsible next step is neither evangelism nor dismissal. It is a research agenda capable of asking better questions. If GLP is overclaimed in some settings, better methods will reveal that. If it has been systematically under-detected because existing designs have privileged analytic legibility, better methods will reveal that too.

The point, in the end, is not to force the field to agree with GLP in advance. It is to insist that disagreement should occur under conditions worthy of the question.

Conclusion: What the Instrument Cannot Yet See

The problem of gestalt language processing in autism research is not simply that the construct remains debated, nor that the available literature is too sparse to permit confident claims. The deeper problem is methodological. Many of the dominant tools used to study autistic language—decontextualised tasks, cross-sectional designs, inherited coding categories, and narrow assumptions about what counts as generative or competent speech—have been calibrated to one developmental logic more than another. Under such conditions, the absence of clear evidence cannot be treated as a neutral finding. It may also

reflect a failure of fit between the architecture being studied and the instrument used to detect it.

That distinction matters because the field has often been too quick to convert surface features into settled explanations. Repetition becomes evidence of non-generativity. Delay becomes evidence of impaired reciprocity. Literalness becomes evidence of inferential deficit. Unusual language becomes evidence of pragmatic weakness. In each case, a behaviour may indeed indicate difficulty—but in each case it may also be doing something more organised, more context-bound, more developmentally dynamic, or more linguistically precise than the inherited category allows. The argument of this paper has not been that GLP should be accepted by assertion, nor that every atypical language feature should be reclassified through a gestalt lens. It has been that the current evidentiary regime has too often resolved the question before it has adequately observed it.

The central recommendation is therefore neither theoretical inflation nor methodological retreat. It is a call for better-fitting evidence: research designs that treat context as part of the construct when appropriate, that follow language across time rather than only in snapshots, that redesign tasks so access conditions can vary without being mistaken for competence itself, and that preserve the distinction between description, function, and developmental inference long enough for claims to remain testable. These are not concessions to subjectivity. They are ordinary requirements of construct validity whenever the phenomenon under study may be organised differently from the assumptions built into the measure.

What the instrument cannot yet see should not be mistaken for what is not there. The task of science, at its best, is not to defend its first tools against the life that exceeds them, but to refine those tools until the life becomes legible on its own terms.

Author Note

No external funding was received for this study. The author declares no conflicts of interest.

References

1. Prizant BM (1982) Gestalt language and gestalt processing in autism. *Topics in Language Disorders* 3: 16–23.
2. Prizant BM (1983) Language acquisition and communicative behavior in autism: Toward an understanding of the “whole” of it. *Journal of Speech and Hearing Disorders* 48: 296–307.
3. Peters AM (1983) Language learning strategies: Does the whole equal the sum of the parts? *Language* 53: 560–573.
4. Volden J, Lord C (1991) Neologisms and idiosyncratic language in autistic speakers. *Journal of autism and developmental disorders* 21: 109–130.
5. Ghaziuddin M, Gerstein L (1996) Pedantic speaking style differentiates Asperger syndrome from high-functioning autism. *Journal of autism and developmental disorders* 26: 585–595.
6. Luyster RJ, Zane E, Wisman Weil L (2022) Conventions for unconventional language: Revisiting a framework for spoken language features in autism. *Autism & developmental language impairments* 7: 23969415221105472.
7. Diehl JJ, Bennetto L, Watson D, Gunlogson C, McDonough J (2008) Resolving ambiguity: a psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and language* 106: 144–152.
8. Whyte EM, Nelson KE (2015) Trajectories of pragmatic and nonliteral language development in children with autism spectrum disorders. *Journal of communication disorders* 54: 2–14.
9. Reindal L, Nærland T, Weidle B, Lydersen S, Andreassen OA, et al. (2023) Structural and Pragmatic Language Impairments in Children Evaluated for Autism Spectrum Disorder (ASD). *Journal of autism and developmental disorders* 53: 701–719.
10. Wu Z, Lam C, To CKS (2025) Spontaneous Generation of Unconventional Language and Its Link with Grammatical Performance in Chinese Adults With and Without ASD. *J Autism Dev Disord* 55: 3539–3552.
11. Zane E, Luyster RJ (2025) An Autistic "Linguatype"? Neologisms, New Words, and New Insights. *Autism research: official journal of the International Society for Autism Research* 18: 1518–1534.
12. Trayvick J, Barkley SB, McGowan A, Srivastava A, Peters AW, et al. (2024) Speech and language patterns in autism: Towards natural language processing as a research and clinical tool. *Psychiatry research* 340: 116109.
13. Hoerricks J (2026) When Delay Is a Category Error: Gestalt Processing and the Misreading of Autistic Development. *J Clin Neuropsychol Prac* 1: 1–10.