Review Article

Beyond ADOS-2-Centered Autism Diagnosis: A Critical Narrative Review of Specificity, Validation Populations, Contemporary Childhood, and Etiological and Functional Assessment

Florencia Maria Lucia Sanabria[1]

1. Instituto HELMAR, Argentina

Background: The Autism Diagnostic Observation Schedule, Second Edition (ADOS-2), can contribute to autism assessment, but its institutional authority can exceed the evidence supporting its use. An observational classification cannot independently establish developmental history, exclude alternative explanations, determine medical etiology, or identify an individual child's treatment priorities.

Objective: To critically examine the justification for ADOS-2-centered diagnostic pathways, emphasizing specificity in clinically complex populations, the independence of validation standards, developmental and contextual measurement bias, examiner competence, and the transportability of historical assessment assumptions to contemporary childhood.

Methods: A critical, nonsystematic narrative synthesis used targeted web searches for PubMed-indexed studies, primary journal records, professional guidance, training documentation, and relevant author-originated conceptual publications. Foundational studies, diagnostic-accuracy syntheses, independent clinical investigations, and context-specific adaptations were distinguished from environmental associations and untested hypotheses. No new patient data or pooled estimates were generated.

Synthesis: Research supports useful average discrimination but does not establish universal specificity or necessity. Pediatric false-positive classifications associated with anxiety and psychiatric complexity, item-level noninvariance associated with language and developmental level, and the need for sensory and age-related adaptations challenge unrestricted transportability. Clinical diagnosis often precedes and remains unchanged after ADOS information, undermining its status as an

obligatory gateway. Digital play, schooling, and nutrition have changed, but these changes have not been shown to invalidate ADOS-2 globally. Their diagnostic relevance requires item-level and prospective investigation. Training attendance, scoring reliability, diagnostic competence, and etiological assessment are distinct responsibilities.

Conclusions: The defensible target of criticism is the elevation of an ADOS-2 classification into a decisive diagnostic, etiological, or service-access verdict. Assessment should organize developmental evidence, differential diagnosis, indicated medical investigation, functional needs, and longitudinal uncertainty before allowing a score to close clinical reasoning. SMART-NB is discussed as an explicitly unvalidated framework for this integration, not as a proven replacement test.

Corresponding author: Florencia Maria Lucia Sanabria, consultorio@draflorenciasanabria.com

1. The clinical problem is the authority assigned to the score

The crucial question is not whether the ADOS-2 records clinically relevant behavior. It is whether a classification produced during a structured encounter should be allowed to organize the whole interpretation of a child's developmental difficulties. When a score becomes the endpoint of reasoning, uncertainty about communication, developmental trajectory, distress, sensory access, and medical contributors can disappear from the report without having been resolved. The resulting document may look standardized while its clinical explanation remains incomplete.

International guidance does not authorize this shortcut. NICE advises against relying on a single autism-specific instrument, and the CDC similarly separates tools from the clinical diagnosis [1][2]. These statements provide a basis for challenging score-determined pathways. They do not establish that every clinician currently misuses the instrument. The present review therefore addresses a practice model and the evidence used to justify it, rather than alleging universal misconduct by ADOS examiners.

Four decisions are commonly conflated: describing observed behavior, determining whether diagnostic criteria are met, investigating contributing or associated medical conditions, and deciding what should be treated first. A measure may help with one decision without validating all four. The clinical literature on autism evaluation and differential diagnosis recognizes the need to integrate developmental, medical, cognitive, linguistic, and psychiatric information [3][4]. The argument developed here is that this

integration should control the interpretation of the observation, rather than being subordinated to its cutoff.

An absence of speech illustrates the problem. Speech production, language comprehension, communicative intention, motor access, social reciprocity, and performance under stress are related but distinguishable functions. A child who does not speak cannot be understood merely by assigning an autism-range observational score. Conversely, speaking fluently does not exclude autism. The meaningful question is what pattern persists across time and contexts after developmental level, access barriers, and plausible alternatives have been considered.

The review takes a deliberately critical position against making ADOS-2 decisive or mandatory. Its conclusions must nevertheless remain conditional on evidence. An instrument can be overvalued without being useless, and a diagnosis can be behaviorally defined without being clinically superficial. Rejecting a test-centered pathway does not justify replacing trained assessment with speculation about hidden biological disease. The standard should be a more accountable clinical formulation, with explicit reasons for both inclusion and exclusion of competing explanations.

2. Review method, evidence hierarchy, and limits

This is a critical narrative review, not a systematic review, diagnostic-accuracy meta-analysis, or validation study. Targeted searches were updated on October 11, 2026. Searches combined ADOS or ADOS-2 with original validation, specificity, false positives, psychiatric disorders, language level, nonverbal mental age, measurement invariance, cultural adaptation, deafness, examiner training, clinical utility, digital play, screen exposure, pandemic development, gastrointestinal symptoms, metabolic investigation, and functional assessment. The search route was web retrieval of indexed bibliographic records, publisher pages, professional guidance, and open repositories; it was not a complete programmatic search of every bibliographic database.

The synthesis prioritizes direct evidence about the instrument over evidence about childhood in general. Foundational and revision studies establish what was developed. Clinical accuracy studies address classification within particular referral populations. Adaptation and invariance studies address specific measurement boundaries. Guidelines address accepted clinical responsibilities. Environmental studies document changes in exposure and developmental context. SMART-NB publications establish the content of a proposed framework; they do not independently validate it. This hierarchy prevents a

biologically plausible mechanism from being promoted into a demonstrated cause of false-positive ADOS classifications.

The transparency principles of SANRA informed the narrative structure [5]. Methodological questions were organized around selection, index-test interpretation, reference-standard independence, and flow or timing, consistent with diagnostic-accuracy appraisal concepts [6][7]. No formal study-by-study QUADAS-2 ratings, duplicate screening, pooled risk-of-bias analysis, protocol registration, or PRISMA flow diagram are claimed. Where the original full text was unavailable, interpretation was restricted to accessible abstracts and publication records. These limitations reduce completeness and must constrain the strength of conclusions.

Because the review deliberately examines arguments against test-centered practice, confirmation bias is a material concern. Favorable accuracy estimates and findings of limited demographic bias are therefore retained alongside critical findings. The issue is not resolved by counting supportive and adverse studies. It requires asking whether each design supports the particular claim being made: useful discrimination, necessity, ecological equivalence, etiological explanation, or improved outcomes. These are different claims with different evidentiary requirements.

3. Historical validation populations and the limits of inheritance

3.1. The original paradigm was population-specific

The 1989 ADOS study described standardized social and communicative observation. Its reliability analyses included matched groups of autistic and nonautistic children with intellectual limitations, aged 6 to 18 years, with a mean IQ of approximately 59 in the specified groups [8]. These were not representative samples of all children presenting to contemporary pediatric services. The historical strength was standardization of an observation that had previously depended heavily on unstructured impressions. Its historical boundary was the population and diagnostic context in which that observation was evaluated.

The ADOS-Generic publication subsequently reported psychometric work involving 223 children and adults and introduced language-based modules [9]. This broadened application but did not turn the sample into a population census. A validation cohort containing several diagnoses can still underrepresent the combinations encountered in ordinary practice: severe language impairment with hearing loss, anxiety with school avoidance, motor impairment with limited access to play, or medically

unstable developmental regression. Generalizability depends on those combinations, not merely on the total sample size.

The 2007 algorithm revision used 1,630 assessments across age and language groupings [10]. The unit was an assessment, so this figure should not automatically be described as 1,630 independent children. Calibrated severity scores and the Toddler Module introduced further developments [11][12]. ADOS-2 therefore cannot accurately be dismissed as an unchanged test from 1989. The stronger criticism is that successive improvements establish validity for specified purposes and populations, not unrestricted equivalence across every new clinical setting.

3.2. Later evidence does not erase sampling boundaries

Historical validation is neither permanent proof nor automatic obsolescence. A procedure may remain useful across decades when the relevant construct and measurement relationship are stable. Equally, a familiar procedure can retain its name while its clinical meaning changes as referral practices, diagnostic criteria, language environments, and case complexity change. Publication date is thus a weak substitute for studying transportability.

The appropriate audit asks which children entered each sample, which were excluded, whether diagnosis was known in advance, how comparison groups were assembled, how often children were assessed, and whether examiners were unusually experienced. It should identify representation by sex, age, language, developmental level, sensory and motor access, psychiatric complexity, cultural context, and geographic setting. Missing reporting is itself an uncertainty, not evidence that a particular group was excluded. No demographic proportions are inferred when the source does not provide them.

A narrowly defined research comparison may distinguish established autism from selected nonspectrum diagnoses well while performing differently among children referred because the diagnosis is uncertain. The latter setting contains competing explanations that resemble the measured phenotype. What must be challenged is the assumption that the authority of a research-derived cutoff transfers intact to this diagnostically ambiguous population.

4. Diagnostic accuracy does not establish a decisive verdict

4.1. High sensitivity cannot substitute for specificity

Lebersfeld and colleagues synthesized 22 studies and reported ADOS-2 sensitivity approximately 0.89-0.92 and specificity 0.81-0.85, while noting limited clinical-setting evidence [13]. The Cochrane review of preschool diagnostic tests found pooled ADOS sensitivity 0.94 and specificity 0.80 across 12 analyses involving 1,625 children, with substantial variation and methodological concerns [14]. The Cochrane estimates combine versions and modules and should not be relabeled as universal ADOS-2 performance. These results support useful discrimination, but not certainty at an individual cutoff.

Fulceri and colleagues' 2025 review rated ADOS-2 evidence more favorably than several alternatives and still located tools within clinical judgment [15]. This favorable evidence matters: simply declaring the test invalid would contradict substantial published research. However, ranking above other tools does not show that the instrument is indispensable, that its specificity is adequate in every referral population, or that using it changes outcomes. Comparative rank, absolute accuracy, and clinical utility are separate questions.

Sensitivity asks how often an autistic participant is classified positively. Specificity asks how often a nonautistic participant is classified negatively. Positive predictive value asks how many positive classifications are correct in the population being tested. A clinic cannot infer the last quantity from sensitivity alone. Reporting that the test is highly sensitive while omitting the referral population and false-positive burden can produce an impression of certainty that the statistical evidence does not support.

4.2. A transparent probability example

Table 1 is an illustrative calculation, not new clinical evidence. Assume sensitivity 90% and specificity 80%, held constant to isolate the effect of pretest probability. Among 1,000 children with a 10% pretest probability, the expected counts are 90 true positives and 180 false positives; only one-third of positive classifications are true positives. With a 50% pretest probability, the corresponding counts are 450 and 100, producing a positive predictive value of approximately 82%. Actual sensitivity and specificity may also change with case mix, making transportability more complicated than this example.

Pretest probability	True positive	False positive	True negative	False negative	Positive predictive value
10%	90	180	720	10	33.3%
30%	270	140	560	30	65.9%
50%	450	100	400	50	81.8%

Table 1. Expected counts per 1,000 children with illustrative sensitivity 90% and specificity 80%

Illustration only. No ADOS module is assigned these parameters. Counts and percentages are expected values, not patient observations.

The calculation exposes an inferential failure, not an estimate that two-thirds of ADOS-positive children are misdiagnosed. The assumed parameters were selected for illustration and are not the performance of a particular module. Its clinical implication is that a threshold must be interpreted within a reasoned estimate of prior likelihood and alternative diagnoses. It cannot independently establish the probability that a given child has autism.

4.3. Direct evidence from complex clinical referrals

Greene and colleagues evaluated 214 patients aged 5-16 years using Module 3. Their abstract reports a 34% false-positive rate and identifies anxiety and trauma-related diagnoses as relevant features. Of 101 patients with elevated scores, 56 were true positives and 45 false positives [16]. Thus, 45/101, approximately 44.6%, is the proportion of positive classifications that were false in that selected sample. It is a different denominator from the reported false-positive rate. Neither figure is a population prevalence of autism overdiagnosis.

This distinction sharpens rather than weakens the criticism. A large fraction of positive observations in a specific complex referral group did not resolve into a clinical autism diagnosis. Anxiety can alter social availability during the encounter, and reduced repetitive-behavior evidence can matter for interpretation. The finding does not show that every anxious child scores positively or that examiner error caused every discrepancy. It shows why nonautism explanations cannot be dismissed after a positive classification.

Medda and colleagues reported variation across modules in a large German clinical sample [17]. Colombi and colleagues examined psychiatric referrals and found limited performance in that context [18]. Adult community mental-health data from Maddox and colleagues provide an additional example of false-positive classifications in psychosis [19]. Adult findings cannot be transferred numerically to children, but the convergence across settings undermines the premise that social-communication differences observed in a brief assessment are inherently specific to autism.

5. Validation bias can make agreement appear more definitive

There is no independent laboratory gold standard for the whole autism spectrum. Clinical expert judgment is therefore a necessary reference, but its construction matters. If the reference diagnosis uses the same ADOS result being evaluated, the index and reference are partly dependent. Agreement may then reflect shared information as well as diagnostic discrimination. This is incorporation bias, not proof that a particular study is fraudulent or that every reported estimate is inflated [6][7].

Selection can create a second problem. Comparing well-characterized autism cases with readily distinguished controls may produce stronger discrimination than comparing diagnostically uncertain referrals. Excluding children who cannot complete a procedure can also narrow the population to which its accuracy applies. The relevant clinical service often sees the excluded or ambiguous children. A validation report must therefore be read through its recruitment and completion pathway before its summary accuracy is applied.

Algorithm development and evaluation within overlapping datasets can create optimism if thresholds are selected to improve separation and then assessed on insufficiently independent data. External validation in a different service, with consecutive referrals and prespecified thresholds, is more informative for transportability. Repeated assessments should be analyzed with attention to clustering within children. A large number of observations does not automatically represent a large number of independent tests of validity.

Observer blinding matters as well. Knowledge of an existing diagnosis, referral narrative, prior test scores, or parental expectations can affect observation or interpretation. Complete blinding may be impractical in clinical work, but validation studies can document which information was available at each stage. Without that documentation, it is difficult to separate the contribution of the observational score from the contribution of the surrounding clinical information.

These concerns do not reverse favorable results automatically. They change the question from whether a paper reports high agreement to whether that agreement is sufficiently independent, representative, and transferable to support the proposed decision. The strongest claim, that ADOS-2 should be decisive, needs more demanding evidence than the modest claim that its observations can be useful.

6. Language and developmental level complicate what the score measures

Zheng and colleagues analyzed Module 1 data from 1,196 children with neurodevelopmental disorders and reported item-level differential functioning associated with spoken language and nonverbal mental age [20]. Their corrigendum corrected specific labels and item descriptions without changing the stated conclusions [21]. This is direct evidence about measurement boundaries. It is more relevant to criticizing universal interpretation than a general statement that contemporary children use different toys.

Noninvariance means that children with comparable underlying trait levels may not have comparable probabilities of particular item responses across groups. Consequently, some score differences can reflect developmental or linguistic characteristics rather than differences in the intended construct alone. This finding concerns the analyzed Module 1 population. It does not establish that all modules are biased in the same way or that every minimally verbal child receives an incorrect classification.

Module selection recognizes expressive-language variation, but it does not resolve the mechanism of that variation. Spoken language can be affected by speech-motor limitations, hearing, cognitive development, anxiety, processing latency, communication opportunities, or neurological disease. The observation should be interpreted alongside these dimensions. Treating language stratification as if it already constitutes differential diagnosis grants the module system an explanatory function it does not possess.

A further clinical danger is equating lack of performance with lack of capacity. A child may understand a task but be unable to execute its motor response, may communicate through an augmentative system that the encounter does not adequately accommodate, or may require a familiar partner to demonstrate a skill. These possibilities are reasons to investigate access and context. They are not reasons to manufacture abilities that have never been demonstrated. Reports should distinguish what was observed, what was not accessible, and what remains unknown.

7. Examiner competence and the misleading language of certification

Kamp-Becker and colleagues studied clinical use of ADOS and ADOS-2, addressing the distinction between research conditions and everyday implementation [22]. Stanford's training documentation explicitly distinguishes introductory instruction, advanced preparation, and demonstration of research reliability; it encourages but does not require research reliability for clinical-only use within its program [23]. This difference provides a concrete basis for questioning what a certificate actually attests to.

Attendance demonstrates exposure to instruction. Administration fidelity demonstrates adherence to a procedure. Scoring reliability demonstrates agreement under specified conditions. Diagnostic competence requires integration of developmental history, criteria, contextual evidence, and alternatives. Etiological assessment additionally requires the appropriate medical expertise and referral relationships. None of these responsibilities can be inferred automatically from another. A person can learn scoring conventions without acquiring the full competence needed to diagnose a complex developmental presentation.

The statement that anyone can apply ADOS after an eight-hour course is not established by the available evidence. Nor does this review establish the absence of every regulatory requirement in Argentina or elsewhere. Professional scope, licensure, publisher qualifications, course requirements, and institutional supervision differ. A credible critique should audit those arrangements rather than assert that they do not exist. The defensible concern is that the word "certified" may conceal substantial differences in the competence actually assessed.

Clinical reports should identify the examiner's professional role, relevant experience, type of training, module selection, accommodations or departures from standardization, and the clinician responsible for diagnosis. Services can separately audit agreement and diagnostic integration. A certificate should not function as a substitute for explaining why an elevated score was accepted, rejected, or held provisionally in a particular case.

Interrater agreement also has limits. Two examiners can agree perfectly on a behavioral code while sharing an incomplete differential formulation. Reliability is necessary for consistent measurement but cannot establish etiological truth. Conversely, a clinician's experience does not excuse careless

administration when a standardized score is reported. Accountability requires both procedural fidelity and explicit recognition of what the procedure cannot decide.

8. Clinical utility challenges obligatory testing

Barbaresi and colleagues' prospective study included 349 young children evaluated by specialist clinicians. Diagnoses before and after ADOS information agreed in 90.0% of cases [24]. The clinicians first completed a developmental and medical evaluation; the study did not compare ADOS with a cursory impression. The final diagnosis incorporated ADOS information and was not an entirely independent biological reference. These details delimit, but do not remove, the relevance of the result.

The study challenges necessity more directly than a specificity estimate does. If a specialist can make a sufficiently confident diagnosis before testing in many cases, making ADOS an obligatory gateway requires an additional justification. That justification might involve difficult or discordant cases, quality assurance, or uncertain presentations. It cannot simply rest on the phrase "gold standard." A test should be selected because it addresses unresolved uncertainty, not because every family must obtain its label before the service acts.

Testing consumes clinician time, family time, and scarce assessment capacity. A pathway can therefore be diagnostically respectable yet inefficient. The opportunity cost is the evaluation or intervention that another child cannot receive while a low-yield assessment is performed. This review does not estimate the size of that cost in Argentina. It identifies the need to evaluate whether mandatory testing improves decisions enough to justify its burden.

The same objection applies to services that treat a negative cutoff as a veto over a well-supported diagnosis or a positive cutoff as a command to diagnose. Both convert a probabilistic measure into administrative authority. Removing that authority should be coupled with a documented assessment standard, accessible second opinions, and support based on identified need. Otherwise, the reform could replace excessive testing with inconsistent judgment rather than improve care.

9. Contemporary childhood and the validity of assessment situations

9.1. Cultural familiarity is empirically relevant

Cepanec and Simlesa examined ADOS-2-defined social-communication behaviors in 220 typically developing Croatian-speaking participants and identified reasons for greater attention to cultural expectations [25]. A deaf adaptation required modification and validation in children and young people [26]. The Adapted ADOS for minimally verbal adolescents and adults adjusted the assessment to an older population [27]. These examples establish that applicability can require more than literal translation or routine module selection.

The adaptations do not prove that the pediatric test is universally outdated. They establish a methodological precedent: when the conditions required to express a skill differ, assessment procedures may need specific validation. Cultural convention, sensory access, developmental expectations, and age appropriateness can affect what an observed response means. A test used in Argentina should therefore be evaluated within relevant linguistic and clinical settings, rather than borrowing credibility solely from international use.

Evidence can also constrain criticism. Kalb and colleagues found minimal overall item-level race and sex bias in a large clinical sample [28]. This finding does not establish equivalence across all cultures or generations, but it prevents a broad assertion that the test is uniformly demographically biased. A strong critique specifies the dimension in which a problem has been demonstrated and the dimension in which it remains a hypothesis.

9.2. Digital childhood raises a testable question

Madigan and colleagues' meta-analysis of 42 studies involving 18,905 children found associations between screen use and language outcomes, with context and content relevant to interpretation [29]. Chu and colleagues reviewed digital play in early childhood [30]. These publications support attention to heterogeneous experiences; they do not demonstrate that screens cause autism or reduce ADOS-2 specificity. Digital exposure may be a correlate, consequence, modifier, or confounded marker of other developmental circumstances.

The more useful question concerns representational familiarity. A child's symbolic competence may be expressed through role play, household routines, narrative games, physical toys, or digital environments. If an assessment elicits only a limited subset of these contexts, failure to demonstrate a particular performance may underrepresent everyday competence. That possibility is an assessment hypothesis. It requires direct comparison within the same children and an independent diagnostic standard.

Replacing materials informally would not answer the question. Once standardized procedures are changed, the validity of their original scoring assumptions may no longer apply. A contemporary-play research protocol should retain the standard assessment, add a separately scored familiar-material condition, and compare classification, item responses, social reciprocity, and generalization. The aim should be equivalence or improved validity, not making all children perform better through easier or coached tasks.

9.3. Schooling pandemic cohorts and nutrition

Kuehn and colleagues reported cohort differences in developmental screening during the pandemic [31]. These data concern screening and particular cohorts, not diagnostic drift of ADOS-2. School attendance, peer opportunities, language environments, and interrupted routines could alter available experience. A clinically useful history should document them, while avoiding the assumption that an atypical social response is necessarily explained by missed schooling.

Wang and colleagues documented changes in ultraprocessed-food consumption among US youths between 1999 and 2018 [32]. That evidence establishes a dietary trend in a specified country. It does not show that changes in food processing caused autism, that they explain today's referral population, or that an ADOS task has become invalid. Nutrition belongs in this discussion as a possible health and performance modifier requiring individual evidence, not as a shortcut from historical exposure to diagnostic causality.

The generational hypothesis is therefore precise: after accounting for developmental level, language, diagnosis, sensory access, socioeconomic circumstances, and examiner behavior, do comparable children from different birth cohorts show different item-response probabilities or diagnostic errors? The relevant analyses include measurement invariance and differential item functioning. The age of a toy alone cannot answer this question, and a null result would constrain the proposed criticism just as a positive result would support it.

10. Organic investigation and the error of premature closure

The ADOS-2 cannot independently identify the cause of a language limitation, a loss of participation, or a distress-related behavior. This is a design boundary. The criticism becomes consequential when clinical services allow the boundary to disappear and present the score as if it has already settled alternatives. Etiological investigation is not an optional philosophical addition when a child's history or examination contains indications that would change management.

Three relationships must remain distinct. A condition may provide a better explanation for the apparent autism-like pattern. It may coexist with autism and independently require treatment. Or it may be associated with the child's autism while leaving the behavioral diagnosis valid. Genetic syndromes, epilepsy, and gastrointestinal disease can occupy different relationships in different children. Finding a condition therefore does not automatically invalidate autism, and failing to identify a molecular cause does not automatically invalidate the diagnosis either [4].

Buie and colleagues' consensus report recognized that gastrointestinal problems can have behavioral manifestations in people with communication difficulties [33]. Sleep assessment also requires attention to contributory conditions and follow-up [34]. The implication for observation is that irritability, withdrawal, agitation, reduced attention, and task refusal may carry information about current distress. It has not been established that each contributor independently increases the ADOS algorithm or that treating it reliably converts a positive score into a negative one.

The Canadian College of Medical Geneticists recommends indication-sensitive genetic and metabolic investigation within neurodevelopmental assessment [35]. Regression, episodic deterioration, unusual neurological findings, syndromic features, and relevant family history can change the investigation pathway. A standardized autism score cannot serve as an exclusion test for these possibilities. The appropriate response is clinically directed investigation, not routine ordering of every metabolic, immune, mitochondrial, gastrointestinal, or inflammatory assay in all children.

Celiac disease illustrates why the critique must remain disciplined. Pediatric neurological manifestations have been reviewed [36], but this does not establish celiac disease as a common cause of absent speech or false-positive ADOS-2 classification. A child with compatible symptoms warrants appropriate evaluation. Improvement in well-being after treatment can be meaningful without proving that the original autism diagnosis was wrong. Temporal improvement alone does not distinguish causation from maturation, concurrent interventions, expectation, or regression to the mean.

A responsible differential report should state the evidence for the developmental phenotype, current medical contributors, investigations indicated by specific findings, and unresolved relationships between them. It should also record why an alternative was considered less likely. This creates a reasoned clinical account that can be revised when new evidence appears. A cutoff does not provide that account.

11. Aggression, irritability, and state-dependent performance

Aggression is not an independent DSM autism criterion, and absence of speech is not a sufficient criterion either [2]. Nevertheless, an encounter can be dominated by refusal, agitation, avoidance, or a striking event. Such behavior can shape the overall clinical impression even when it is not itself a scored diagnostic criterion. Reports must distinguish a descriptive event from the trait or condition inferred from it.

Not every child who hits has a stable aggressive disposition. Hitting can occur during pain, fear, communication breakdown, overload, frustration, or deliberate interpersonal aggression. These possibilities require different formulations and cannot be distinguished by moralizing the behavior. Equally, the reverse claim that aggressive behavior in neurodevelopment is always irritability or dysregulation is unsupported. The task is to test functions and contributors rather than rename all behavior in one direction.

A useful performance model separates demonstrated capacity, access to the task, current physiological and emotional state, environmental demand, and the support required. This model is a clinical heuristic, not a validated equation. It explains why an observation obtained under one condition should not automatically be treated as the child's stable maximum or everyday baseline. Repetition across appropriate settings can show whether the limitation is persistent, contingent, or still unexplained.

The distinction is especially important for limited spoken communication. A child may signal pain through changes in sleep, movement, eating, or participation. A clinician should investigate the temporal pattern and associated signs. Improvement in distress may justify revising the functional plan while leaving the autism diagnosis unchanged. Diagnostic formulation and functional monitoring are related tasks, but they should not be forced to move together every time a symptom changes.

12. Possible harms must be studied rather than presumed

False-positive classification, incorrect clinical diagnosis, delayed recognition of a coexisting condition, and inappropriate treatment are distinct outcomes. An elevated score is not itself proof that the child received an incorrect diagnosis, because a competent team may interpret it appropriately. Likewise, a later diagnosis does not establish that the ADOS caused its omission. Causal claims require documentation of the decision pathway, relevant timing, and alternative explanations.

Potential consequences of premature closure include misdirected interventions, unnecessary restrictions, missed medical treatment, school decisions based on an incomplete profile, and family expectations shaped by excessive certainty. These are plausible clinical risks, not prevalence estimates demonstrated by this review. The benefit of a correct autism diagnosis, including understanding, accommodations, and support, must also be recognized. The target is diagnostic error and unjustified authority, not autistic identity or the child's right to assistance.

An indiscriminate biological pathway can generate harms of its own: false-positive laboratory results, burdensome investigations, dietary inadequacy, unsupported causal explanations, and delays in needed developmental support. The criticism of ADOS-centered care becomes credible only if the proposed replacement process accounts for these risks. More testing is not synonymous with better diagnosis. The clinically relevant endpoint is a more useful and accurate decision with proportionate burden.

Future studies should therefore measure management changes, missed actionable conditions, unnecessary testing, time to support, family burden, functional outcomes, and diagnostic stability. They should distinguish harms of the instrument from harms of its interpretation or institutional use. A service-level reform should be judged by these outcomes rather than by the number of scores or biological explanations it produces.

13. SMART NB as an integration framework with an explicit evidence boundary

13.1. Existing author publications and the proposed extension

The SMART-NB white paper describes a longitudinal framework for translating clinical and functional information into intervention priorities [37]. Related conceptual publications address communication and language, intellectual disability, and childhood anxiety [38][39][40]. Sanabria's nutrition and energy-

metabolism review situates directed organic evaluation across ten operational clinical profiles [41]. These publications are cited for conceptual provenance. Repository publication and DOI assignment are not demonstrations of diagnostic validity or superiority.

The white paper explicitly positions SMART-NB as a layer of integration and follow-up that does not replace established diagnosis. This boundary must be preserved when connecting it to criticism of ADOS-2. The proposal is to replace the organization of a score-centered pathway, not to declare SMART-NB a validated autism classifier. Calling the framework a proven replacement test would contradict the status of the author's own source documents.

Its ten operational profiles include autism, ADHD, communication/language, intellectual disability, motor disorders, learning disorders, anxiety, depression, psychosis, and oppositional-defiant presentations. This is a cross-diagnostic clinical organization; it is not a claim that all ten belong to the DSM neurodevelopmental-disorders chapter. Profiles can overlap, and the categorical diagnosis remains a separate clinical responsibility.

13.2. Seven functional domains and longitudinal priorities

The seven white-paper domains are communication/language, learning/problem solving, attention/executive functions, adaptive autonomy, social participation, motor function/praxis, and regulation/self-management [37]. Their purpose is to describe the child's current functioning and relevant supports. A profile can reveal differences hidden by a single total. Two children with the same observational classification may need entirely different first interventions.

The framework is conceptually compatible with the WHO approach to functioning and participation [42]. PEDI-CAT and goal-attainment scaling provide established references for functional measurement and individualized goals [43][44]. Compatibility does not transfer their validation to SMART-NB. The proposed Global Functional Index requires its own testing of content validity, reliability, missing-data rules, interpretability, and responsiveness. An equal total can conceal improvement in one domain and deterioration in another.

A practical integration sequence asks what was measured, which contexts modify its expression, what hypotheses are supported, which needs take priority, what action is justified, and how outcome will be reassessed. The method should make uncertainty visible. For example, unavailable hearing information

should remain unknown rather than being coded as normal; a suspected organic contributor should remain a hypothesis until appropriate evaluation supports it.

13.3. High and low demand require operational clarification

The phrase demand has two different uses that should not be collapsed. Environmental demand concerns the complexity or load of a task. Clinical burden concerns the child's current pain, fatigue, irritability, distress, or reduced functional availability. A low-demand environment can contain a child with high clinical burden. A demanding task can reveal limitations in a child who is medically comfortable. Using one label for both would weaken reproducibility.

For this proposal, a high clinical-burden profile describes substantial distress or effort associated with participation, while a low-output profile describes reduced initiation, response, or execution. These are descriptive hypotheses, not etiological diagnoses, and they may coexist. Both should be examined under separately recorded environmental demand. A quiet child must not be presumed to have low distress, and a highly active child must not be presumed to have adequate capacity.

This operational clarification extends the integration proposal and is not presented as an already validated SMART-NB scale. Research should specify observable indicators, duration, context, supports, recovery, and alternative explanations before using the categories. No sensitivity, specificity, thresholds, normative values, or validated treatment rules are assigned here. Their usefulness will depend on whether independent clinicians can apply them consistently and whether they improve decisions beyond existing good clinical practice.

14. Clinical illustrations show decision differences without inventing evidence

The following scenarios are hypothetical and contain no patient records. In the first, a four-year-old has limited speech, disrupted sleep, bowel discomfort, and reduced participation during an unfamiliar assessment. A positive observation does not identify which components reflect autism, distress, or both. The proposed pathway documents developmental evidence, communication access, indicated evaluation of the discomfort, and functional supports. It does not postpone all assistance until every investigation is complete.

In the second, a verbally fluent child speaks freely at home but becomes avoidant in evaluative settings. A score elevated by social unavailability must be interpreted against onset, repetitive-behavior history, school context, anxiety, and consistency across situations. Neither good home language nor situational anxiety independently settles the diagnosis. The relevant change is that an alternative explanation remains actively testable after the positive observation.

In the third, a child with global developmental limitations cannot complete a play routine requiring motor planning and unfamiliar representational knowledge. The examiner should distinguish developmental expectation, physical access, familiarity, and the autism-related pattern. A score cannot establish that all failures belong to one condition. Repeated accessible observation can improve the functional formulation without informally modifying a standardized score and claiming unchanged validity.

In the fourth, a child loses previously acquired language and participation. The temporal change makes neurological assessment clinically important regardless of the ADOS classification. The research question is not whether the test contains a regression diagnosis, but whether the care pathway recognizes the information that the test cannot decide. In each scenario, improvement after treatment would support monitoring and reconsideration; it would not alone prove an original diagnostic error.

15. A research agenda that can actually challenge the instrument

A study of contemporary transportability should recruit consecutive referrals and specify age, language, developmental level, sensory or motor differences, psychiatric complexity, medical burden, cultural setting, and birth cohort. A contemporaneous older group is not by itself a historical cohort comparison: age and generation are confounded. Designs need archived comparable data or appropriately matched repeated cohorts, with changes in referral thresholds and diagnostic practice considered explicitly.

The reference diagnosis should use an independent multidisciplinary process with the ADOS score concealed during initial adjudication when feasible. Disagreement should be reviewed rather than resolved automatically in favor of the test. Both false positives and false negatives require follow-up. The study should retain children who cannot complete standard testing in the service-level analysis, even if a particular measurement analysis cannot include them.

The play hypothesis requires item-level analysis of familiarity, accessible performance, and diagnostic errors rather than a comparison of total scores alone. Prespecified measures should examine physical and

digital play separately, caregiver participation, language exposure, schooling, and task novelty. Nutrition and medical burden should be analyzed as measured clinical characteristics, not broad proxies for a modern generation. Associations must not be presented as mechanisms without additional evidence.

SMART-NB requires a different evaluation because it is a complex clinical process with proposed measurement components. Development, feasibility, and controlled evaluation should follow principles appropriate to complex interventions [45]. Initial studies should test whether clinicians agree on functional priorities and burden/output descriptors. Subsequent studies should compare outcomes against a strong standard-care pathway, not against deliberately inadequate ADOS-only care. Independent investigators and prespecified analyses are particularly important for a framework developed by the reviewing author.

Primary outcomes might include appropriate identification of actionable coexisting conditions, independently judged management changes, time to needed support, and improvement in meaningful participation. Safety outcomes should include unnecessary investigation, dietary restriction, missed autism, family burden, and inequitable access. Any patient-data study requires appropriate ethical review and consent arrangements before collection or secondary use. No such approval or study result is implied by this conceptual review.

15.1. Even recent pooled evidence requires an extraction audit

A 2026 module-specific synthesis reports pooled sensitivity 0.88 and specificity 0.74 [46]. Its published text reconstructs the Greene comparison as 55 true positives, 34 false positives, one false negative, and 124 true negatives. The original abstract instead reports 56 true positives and 45 false positives among 101 elevated results [16]. The reconstructed table also implies a false-positive rate of 34/(34+124), approximately 21.5%, rather than the original reported 34%.

This is an apparent extraction discrepancy that needs clarification, not a recalculated meta-analysis or a finding of misconduct. The present review therefore does not use that newer pooled result as its principal quantitative anchor. The discrepancy demonstrates why a critical paper should verify primary numerators and denominators before repeating a headline estimate. Publication recency and sophisticated modeling do not exempt a synthesis from data auditing.

16. Objections that a strong critique must answer

One objection is that ADOS was never intended to determine etiology. That is correct. The criticism is directed at the clinical authority assigned to its output, not at the absence of a function the developers did not claim. The relevant empirical question is whether test-centered pathways suppress differential reasoning, and whether reorganizing them improves outcomes. This remains a research question where direct service-level evidence is incomplete.

A second objection is that expert judgment can also be biased. That is also correct. The alternative must therefore be structured, documented, and open to external review. Replacing one instrument's authority with one clinician's unchallengeable authority would not improve the evidentiary standard. The proposal requires explicit hypotheses, criteria-based diagnosis, independent measurement where appropriate, and opportunities for reassessment.

A third objection is that rejecting mandatory testing may reduce diagnostic consistency. Consistency should be protected by common reporting requirements, training, independent audits, and selective use of validated instruments. The question is whether universal administration is the best route to that goal. A targeted instrument strategy can preserve useful observation while directing scarce resources toward cases in which it resolves genuine uncertainty.

A fourth objection concerns overlap with existing models of comprehensive assessment. SMART-NB cannot claim that medical history, context, participation, or longitudinal follow-up are newly discovered priorities. Its potential contribution lies in reproducibly connecting these priorities to decisions and outcomes. Its own conceptual sources make that ambition testable [37][38][39][40][41]. Until tested, it is a proposal with a development agenda, not evidence that the author has solved the limitations identified in ADOS-2.

17. Conclusions

The strongest evidence-based criticism of ADOS-2 is that its useful observational discrimination has been granted a degree of diagnostic and institutional authority that its design and validation cannot universally sustain. Specificity varies with clinical complexity; developmental and linguistic differences affect some measurement relationships; training does not automatically establish diagnostic competence; and clinical utility does not justify making the procedure obligatory in every presentation.

Historical populations and contemporary childhood strengthen the case for reassessing transportability, but they do not establish that traditional play materials have become diagnostically invalid. The hypothesis must be tested at the level of items, accessible performance, and independent diagnosis. Likewise, organic conditions warrant investigation when clinically indicated, but neither their discovery nor improvement after treatment automatically negates autism.

An ADOS-2 classification should have no independent authority to establish etiology, end differential diagnosis, determine a child's full functional needs, or act as a universal gate to support. Its clinical value must be justified case by case within a broader formulation. SMART-NB offers a proposed way to organize that formulation and follow-up. Its usefulness should be demonstrated under the same critical standards demanded of established instruments.

Statements and Declarations

Funding

No specific funding was received for this work.

Potential Competing Interests

The author developed SMART-NB, directs Instituto HELMAR, and is professionally involved in clinical practice and educational activities associated with the framework. These activities create a relevant intellectual and professional interest in its evaluation. No financial benefit or independent validation is inferred from this review.

Ethics

This article reviews published literature and conceptual documents. It reports no new participant research or identifiable patient information; the clinical scenarios are hypothetical. No institutional ethics approval is claimed.

Data Availability

No new patient dataset was generated or analyzed. The illustrative probability table is fully specified in the text. Sources are identified in the reference list.

Author Contributions

Florencia Maria Lucia Sanabria developed the clinical question and SMART-NB conceptual proposal. The manuscript incorporates AI-assisted literature location, synthesis, drafting, translation, and organization. Final source verification, interpretation, and approval remain the author's responsibility.

Use of Generative AI

ChatGPT was used to assist with literature searches, critical organization, scientific drafting, and preparation of the English and Spanish versions. AI is not an author. The submitting author is responsible for reviewing the entire manuscript, checking source accuracy, and approving its conclusions before submission.

References

1. ^*National Institute for Health and Care Excellence (2011). "Autism Spectrum Disorder in Under 19s: Recognition, Referral and Diagnosis." CG128. National Institute for Health and Care Excellence. https://www.nice.org.uk/guidance/cg128/chapter/Recommendations.*
2. a, b*Centers for Disease Control and Prevention (2025). "Clinical Testing and Diagnosis for Autism Spectrum Disorder." Centers for Disease Control and Prevention. https://www.cdc.gov/autism/hcp/diagnosis/index.html.*
3. ^*Hyman SL, Levy SE, Myers SM (2020). "Identification, Evaluation, and Management of Children With Autism Spectrum Disorder." Pediatrics. **145**:e20193447. doi:10.1542/peds.2019-3447.*
4. a, b*Olson L, Bishop S, Thurm A (2024). "Differential Diagnosis of Autism and Other Neurodevelopmental Disorders." Pediatr Clin North Am. **71**:157–177. doi:10.1016/j.pcl.2023.12.004.*
5. ^*Baethge C, Goldbeck-Wood S, Mertens S (2019). "SANRA—a Scale for the Quality Assessment of Narrative Review Articles." Res Integr Peer Rev. **4**:5. doi:10.1186/s41073-019-0064-8.*
6. a, b*Whiting PF, Rutjes AWS, Westwood ME, et al. (2011). "QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies." Ann Intern Med. **155**:529–536. doi:10.7326/0003-4819-155-8-201110180-00009.*
7. a, b*Bossuyt PM, Reitsma JB, Bruns DE, et al. (2015). "STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies." BMJ. **351**:h5527. doi:10.1136/bmj.h5527.*

8. ^ Lord C, Rutter M, Goode S, et al. (1989). "Autism Diagnostic Observation Schedule: A Standardized Observation of Communicative and Social Behavior." J Autism Dev Disord. **19**:185–212. doi:10.1007/BF02211841.
9. ^ Lord C, Risi S, Lambrecht L, et al. (2000). "The Autism Diagnostic Observation Schedule-Generic: A Standard Measure of Social and Communication Deficits Associated With the Spectrum of Autism." J Autism Dev Disord. **30**:205–223. doi:10.1023/A:1005592401947.
10. ^ Gotham K, Risi S, Pickles A, Lord C (2007). "The Autism Diagnostic Observation Schedule: Revised Algorithms for Improved Diagnostic Validity." J Autism Dev Disord. **37**:613–627. doi:10.1007/s10803-006-0280-1.
11. ^ Gotham K, Pickles A, Lord C (2009). "Standardizing ADOS Scores for a Measure of Severity in Autism Spectrum Disorders." J Autism Dev Disord. **39**:693–705. doi:10.1007/s10803-008-0674-3.
12. ^ Luyster R, Gotham K, Guthrie W, et al. (2009). "The Autism Diagnostic Observation Schedule-Toddler Module: A New Module of a Standardized Diagnostic Measure for Autism Spectrum Disorders." J Autism Dev Disord. **39**:1305–1320. doi:10.1007/s10803-009-0746-z.
13. ^ Lebersfeld JB, Swanson M, Clesi CD, O'Kelley SE (2021). "Systematic Review and Meta-Analysis of the Clinical Utility of the ADOS-2 and the ADI-R in Diagnosing Autism Spectrum Disorders in Children." J Autism Dev Disord. **51**:4101–4114. doi:10.1007/s10803-020-04839-z.
14. ^ Randall M, Egberts KJ, Samtani A, et al. (2018). "Diagnostic Tests for Autism Spectrum Disorder (ASD) in Preschool Children." Cochrane Database Syst Rev. **7**:CD009044. doi:10.1002/14651858.CD009044.pub2.
15. ^ Fulceri F, Caruso A, Micai M, et al. (2025). "Autism Diagnosis in Children and Adolescents: A Systematic Review and Meta-Analysis of Test Accuracy." Neurosci Biobehav Rev. **173**:106164. doi:10.1016/j.neubiorev.2025.106164.
16. ^ a, b Greene RK, Vasile I, Bradbury KR, Olsen A, Duvall SW (2022). "Autism Diagnostic Observation Schedule (ADOS-2) Elevations in a Clinical Sample of Children and Adolescents Who Do Not Have Autism: Phenotypic Profiles of False Positives." Clin Neuropsychol. **36**:943–959. doi:10.1080/13854046.2021.1942220.
17. ^ Medda JE, Cholemkery H, Freitag CM (2019). "Sensitivity and Specificity of the ADOS-2 Algorithm in a Large German Sample." J Autism Dev Disord. **49**:750–761. doi:10.1007/s10803-018-3750-3.
18. ^ Colombi C, Fish A, Ghaziuddin M (2020). "Utility of the ADOS-2 in Children With Psychiatric Disorders." Eur Child Adolesc Psychiatry. **29**:989–992. PMID 31587085.
19. ^ Maddox BB, Brodkin ES, Calkins ME, et al. (2017). "The Accuracy of the ADOS-2 in Identifying Autism Among Adults With Complex Psychiatric Conditions." J Autism Dev Disord. **47**:2703–2709. doi:10.1007/s10803-017-3188-z.

20. ^ Zheng S, Kaat A, Farmer C, et al. (2022). "Bias in Measurement of Autism Symptoms by Spoken Language Level and Non-Verbal Mental Age in Minimally Verbal Children With Neurodevelopmental Disorders." Front Psychol. **13**:927847. doi:10.3389/fpsyg.2022.927847.

21. ^ Zheng S, Kaat A, Farmer C, et al. (2022). "Corrigendum: Bias in Measurement of Autism Symptoms by Spoken Language Level and Non-Verbal Mental Age in Minimally Verbal Children With Neurodevelopmental Disorders." Front Psychol. **13**:1051464. doi:10.3389/fpsyg.2022.1051464.

22. ^ Kamp-Becker I, Albertowski K, Becker J, et al. (2018). "Diagnostic Accuracy of the ADOS and ADOS-2 in Clinical Practice." Eur Child Adolesc Psychiatry. doi:10.1007/s00787-018-1143-y.

23. ^ Stanford Medicine Autism Center (2026). "ADOS-2 Training FAQs." Stanford Medicine Autism Center. https://med.stanford.edu/autismcenter/autism-diagnostic-training-program-/ados-2-training-faqs.html.

24. ^ Barbaresi W, Cacia J, Friedman S, et al. (2022). "Clinician Diagnostic Certainty and the Role of the Autism Diagnostic Observation Schedule in Autism Spectrum Disorder Diagnosis in Young Children." JAMA Pediatr. **176**:1233–1241. doi:10.1001/jamapediatrics.2022.3605.

25. ^ Cepanec M, Simlesa S (2024). "Item-Based Analysis of Some ADOS-2 Items With Typically Developing Participants Might Help Improve Cross-Cultural Validity of ADOS-2." J Autism Dev Disord. **54**:109–120. doi:10.1007/s10803-022-05791-w.

26. ^ Phillips H, Wright B, Allgar V, et al. (2022). "Adapting and Validating the Autism Diagnostic Observation Schedule Version 2 for Use With Deaf Children and Young People." J Autism Dev Disord. **52**:553–568. doi:10.1007/s10803-021-04931-y.

27. ^ Bal VH, Maye M, Salzman E, et al. (2020). "The Adapted ADOS: A New Module Set for the Assessment of Minimally Verbal Adolescents and Adults." J Autism Dev Disord. **50**:719–729. doi:10.1007/s10803-019-04302-8.

28. ^ Kalb LG, Singh V, Hong JS, et al. (2022). "Analysis of Race and Sex Bias in the Autism Diagnostic Observation Schedule (ADOS-2)." JAMA Netw Open. **5**:e229498. doi:10.1001/jamanetworkopen.2022.9498.

29. ^ Madigan S, McArthur BA, Anhorn C, Eirich R, Christakis DA (2020). "Associations Between Screen Use and Child Language Skills: A Systematic Review and Meta-Analysis." JAMA Pediatr. **174**:665–675. doi:10.1001/jamapediatrics.2020.0327.

30. ^ Chu C, Paatsch L, Kervin L, Edwards S (2024). "Digital Play in the Early Years: A Systematic Review." Int J Child-Comput Interact. **40**:100652. doi:10.1016/j.ijcci.2024.100652.

31. ^ Kuehn LM, Jones A, Helmkamp L, Knudtson M, Domek GJ, Allison MA (2024). "Socioemotional Development of Infants and Toddlers During the COVID-19 Pandemic." JAMA Pediatr. **178**:151–159. doi:10.1001/jamapediatrics.2023.5684.

32. ^ Wang L, Martinez Steele E, Du M, et al. (2021). "Trends in Consumption of Ultraprocessed Foods Among US Youths Aged 2-19 Years, 1999-2018." JAMA. **326**:519–530. doi:10.1001/jama.2021.10238.

33. ^ Buie T, Campbell DB, Fuchs GJ, et al. (2010). "Evaluation, Diagnosis, and Treatment of Gastrointestinal Disorders in Individuals With ASDs: A Consensus Report." Pediatrics. **125**(Suppl 1):S1–S18. doi:10.1542/peds.2009-1878C.

34. ^ Malow BA, Byars K, Johnson K, et al. (2012). "A Practice Pathway for the Identification, Evaluation, and Management of Insomnia in Children and Adolescents With Autism Spectrum Disorders." Pediatrics. **130**(Suppl 2):S106–S124. PMID 23118242.

35. ^ Carter MT, Srour M, Au PYB, et al. (2023). "Genetic and Metabolic Investigations for Neurodevelopmental Disorders: Position Statement of the Canadian College of Medical Geneticists (CCMG)." J Med Genet. **60**:523–532. doi:10.1136/jmg-2022-108962.

36. ^ Lionetti E, Francavilla R, Pavone P, et al. (2010). "The Neurology of Coeliac Disease in Childhood: What is the Evidence? A Systematic Review and Meta-Analysis." Dev Med Child Neurol. **52**:700–707. doi:10.1111/j.1469-8749.2010.03647.x.

37. a, b, c Sanabria F (2026). "SMART-NB White Paper v1.0. Del diagnóstico a las prioridades funcionales longitudinales en neurodesarrollo pediátrico" [SMART-NB White Paper v1.0. From Diagnosis to Longitudinal Functional Priorities in Pediatric Neurodevelopment]. HELMAR. Conceptual and methodological specification, pre-validation. doi:10.5281/zenodo.22979890.

38. a, b Sanabria FML (2026). "Beyond the DSM-5-TR Communication and Language Profile: A Functional Reinterpretation Through the SMART-NB Methodology." HELMAR. Conceptual repository manuscript, version 2. doi:10.5281/zenodo.22994009.

39. a, b Sanabria FML (2026). "Beyond the DSM-5-TR Intellectual Disability Profile: A Functional Reinterpretation Through the SMART-NB Methodology." HELMAR. Conceptual repository manuscript. doi:10.5281/zenodo.22994749.

40. a, b Sanabria FML (2026). "Beyond the DSM-5-TR Child and Adolescent Anxiety Profile: A Functional Reinterpretation Through the SMART-NB Methodology." HELMAR. Conceptual repository manuscript, version 2. doi:10.5281/zenodo.22999541.

41. a, b Sanabria FML (2026). "Nutrição e metabolismo energético no diagnóstico diferencial de dez perfis clínicos do neurodesenvolvimento" [Nutrition and Energy Metabolism in the Differential Diagnosis of Ten Clinical Neurodevelopmental Profiles]. SciELO Preprints. Preprint, not peer reviewed. Public-record title; Spanish

PDF entitled Nutrición y metabolismo energético en el diagnóstico diferencial de diez perfiles clínicos del ne urodesarrollo. doi:10.1590/SciELOPreprints.18348.

42. ^*World Health Organization (2007). International Classification of Functioning, Disability and Health: Chil dren and Youth Version. Geneva: WHO. https://www.who.int/news/item/23-10-2007-who-publishes-new-st andard-for-documenting-the-health-of-children-and-youth.*
43. ^*Haley SM, Coster WJ, Dumas HM, et al. (2011). "Accuracy and Precision of the Pediatric Evaluation of Disab ility Inventory Computer-Adaptive Tests (PEDI-CAT)." Dev Med Child Neurol. **53**:1100–1106. doi:10.1111/j.146 9-8749.2011.04107.x.*
44. ^*Kiresuk TJ, Sherman RE (1968). "Goal Attainment Scaling: A General Method for Evaluating Comprehensiv e Community Mental Health Programs." Community Ment Health J. **4**:443–453. doi:10.1007/BF01530764.*
45. ^*Skivington K, Matthews L, Simpson SA, et al. (2021). "A New Framework for Developing and Evaluating Co mplex Interventions: Update of Medical Research Council Guidance." BMJ. **374**:n2061. doi:10.1136/bmj.n206 1.*
46. ^*Kilincel S, Bulut F, Goksel P, Usta MB, Kilincel O (2026). "Module-Specific Diagnostic Accuracy of ADOS-2 in Real-World Clinical Referral Populations: An Updated Systematic Review and HSROC Meta-Analysis." Front Psychiatry. **17**:1840734. doi:10.3389/fpsyt.2026.1840734.*

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: The author developed SMART-NB, directs Instituto HELMAR, and is professionally involved in clinical practice and educational activities associated with the framework. These activities create a relevant intellectual and professional interest in its evaluation. No financial benefit or independent validation is inferred from this review.