DOCUMENT SUMMARY

This article by Dr. Laurent Mottron argues that autism research has stagnated due to the over-inclusivity of current diagnostic criteria and an over-reliance on standardized instruments like the ADOS and ADI-R. Mottron posits that this has led to increasingly heterogeneous research cohorts, which masks real findings and hinders scientific progress. This document is critically important for Enlitens as it provides a robust, peer-reviewed argument from a leading researcher against the validity of the very standardized tools Enlitens seeks to replace, while advocating for a return to expert clinical judgment based on "prototypes," a methodology that aligns perfectly with the Enlitens Clinical Interview.

FILENAME

MOTTRON_2021_Autism_Research_Strategy_critique_of_standardized_instruments_and_support_for_prototype_assessment.md

METADATA

- **Primary Category**: ASSESSMENT
- **Document Type**: research_article
- **Relevance**: Core
- **Key Topics**: assessment_critique, standardized_testing, clinical_judgment, neurodiversity, autism, diagnostic_criteria, validity, reliability
- **Tags**: #autism, #assessment, #critique, #standardized_testing, #ADOS, #ADI-R, #clinical_interview, #prototype, #heterogeneity, #neurodiversity

CRITICAL QUOTES FOR ENLITENS

- "In research, the inclusion of individuals categorically defined by over-inclusive, polythetic criteria in autism cohorts results in a population whose heterogeneity runs contrary to the advancement of scientific progress."
- "The diagnosis of autism is obtained using these instruments when reaching a threshold summary score by adding individual item scores... Their cut-off threshold scores are determined by a specificity-sensitivity trade-off, expert agreement long ago being their reference."
- "Multiple warnings, especially by C. Lord, that they should not be used alone and without a clinical judgment have been essentially abolished by their commercial presentation as diagnostic instruments."
- "Autism in the clinical and research world of today is what is measured by the ADI-R and ADOS-G and reliability is confused with truth."

- "Overall, polythetic criteria and their ascension in a hierarchical classification increase reliability, but at the risk of turning it into triviality: if judges are divided to decide whether a bumpy circle is indeed a circle, they will all agree that both are shapes."
- "The abstract nature of certain DSM 5 criteria of autism (e.g., A3: deficits in developing and maintaining relationships) is a dramatic example of reliability turned into triviality."
- "The autism spectrum as currently defined by the DSM and operationalized by standardized tools should not be the starting point for scientific research in neuroscience."
- "The choice of considering dimensions to be relevant is even more arbitrary than that of the limits of categories."

KEY STATISTICS & EVIDENCE

- Recent meta-analytical studies indicate that case-control effect sizes have decreased by up to 80% for neurocognitive constructs (emotional recognition, planning, capacity of cognitive perspective taking, brain size, and EEG characteristics) that distinguish autistic from non-autistic people.
- There has been a gradual 30-fold increase in the prevalence of people diagnosed as autistics over the last 50 years, which coincides with the inclusion of individuals who are increasingly distant from the initial description.
- The number of signs required to provide an autism diagnosis decreased by a factor of two between 2004-2005 and 2014 for children diagnosed at school age in Sweden.

METHODOLOGY DESCRIPTIONS

Critique of Standardized Assessment Methodologies

The author identifies several methodological dogmas that have led to the current crisis in autism research, directly challenging the principles of standardized assessment.

Methodological dogmas, premature assumptions, and case ascertainment strategies that contribute to the trivialisation of autism

Reliability/standardization: The dogma for the diagnosis of autism is the use of validated and standardized instruments that unify the operationalization of DSM criteria and reduce the discrepancy between individual judgments. We suspect such standardisation of diagnostic procedures to be largely responsible for the plateauing of autism research, by adding artifactual or criteria- or instrument-based heterogeneity to the natural variability of autistic presentation due to sex, age and outcome. The diagnosis of autism is obtained using these instruments when reaching a threshold summary score by adding individual item scores (Randall et al., 2018). Their cut-off threshold scores are determined by a specificity-sensitivity trade-off, expert agreement long ago being their reference. Multiple warnings, especially by C. Lord, that they should not be used alone and without a clinical judgment have been essentially abolished by their commercial presentation as diagnostic instruments. However, we now know that such instruments are over-inclusive (Molloy, Murray, Akers, Mitchell, & Manning-Courtney, 2011), influenced by non-specific dimensions (Fombonne et al., 2020; Havdahl et al., 2016), and vulnerable to large-scale temporal evolution (Arvidsson et al., 2018). Despite such warnings,

most research papers use them as an entry point without further refinement. Autism in the clinical and research world of today is what is measured by the ADI-R and ADOS-G and reliability is confused with truth.

The issue with standardized instruments may be intrinsic to the use of summary scores of polythetic criteria. Summary scores privilege the grouping of exemplars that share certain features -trivial when quantitatively measured- over the intersection of maximally resembling exemplars. Moreover, signs in standardized instruments are independent to avoid a "halo effect", that is the bias to detect one sign when another related one is present - the negative counterpart of expertise. Therefore, their grouping into "metasigns", subsets of qualitatively specified signs that strengthen the clinical recognition of the diagnosis when present together, is lost in the operation. Furthermore, signs in polythetic systems are not differentially weighted: their contribution to the varying distance from the prototype is replaced by a global quantitative pass or fail. Overall, polythetic criteria and their ascension in a hierarchical classification increase reliability, but at the risk of turning it into triviality: if judges are divided to decide whether a bumpy circle is indeed a circle, they will all agree that both are shapes. The abstract nature of certain DSM 5 criteria of autism (e.g., A3: deficits in developing and maintaining relationships) is a dramatic example of reliability turned into triviality.

Sample size. A conviction shared by the scientific community in autism is that the first research on small samples biased the results in favor of their initial hypotheses, whereas studies on a large N, with high standards, brought the previously found results into a more just light. This belief is consistent with the belief that meta-analyses provide us with a safer message than individual studies. However, there are undeniable examples (e.g. in intervention: Pickles et al., 2016) in which a single study is better than a thousand studies with lower standards (Dawson & Fletcher-Watson, 2020). Moreover, in the current state of the definition of the autism spectrum, the primacy attributed to the size of the sample over the resemblance of the individuals who compose it creates a level of noise that increases dramatically with the size of the sample. The avoidance of the type-1 risk associated with small samples must be balanced against the type-2 risk associated with large, heterogeneous samples.

Representativeness. Ensuring the representativeness of the sample tested for the population under study will always prompt us to favor probability sampling over convenience sampling. However, random sampling within large cohorts is obtained at the cost of a constant rise in the hierarchical taxonomy of neurodevelopmental conditions, with which autism then becomes confused: any neurodevelopmental or adult psychiatric condition is now suspected to have autistic traits. Probability sampling only makes sense for a population for which the identification is unquestionable and can be taken as a starting point for research.

A Proposed Alternative Methodology for Assessment

The author proposes a specific, multi-step process for creating research cohorts that prioritizes clinical judgment and prototypicality over standardized scores. This method can serve as a model for Enlitens' own assessment philosophy.

We therefore propose the following steps for the creation of a research cohort combining the advantages of standardized categorical type diagnosis and gradation in prototypicality.

a) Sample a population that exceeds the sum score of a standardized threshold.

b) Decompose the population into compartments with homogeneous values for the DSM 5 specifiers (e.g., comorbidity: with vs without CNV or neuro-genetic conditions; language: with vs without initial language delay; intelligence: with vs without non-verbal intellectual disability) to which will be added age (preschool vs. school and adult age) and sex.

c) Classify in situ individuals who make up these compartments by decreasing prototypicality. This ranking is obtained by averaging the score of each participant according to two experts based on the following elements: level of similarity to his personal autism category, speed of clinical identification, exemplarity for academic teaching.

d) Determine an N sufficient for the desired power and truncate the compartments to these Ns.

e) Finally, compare the case-control differences obtained in each of these compartments to test their generalizability.

THEORETICAL FRAMEWORKS

Prototype Theory as an Alternative to Standardized Cut-offs

The author provides a strong theoretical basis for prioritizing clinical judgment over algorithmic scoring by invoking prototype theory from cognitive psychology.

Contribution of prototype theory

Reaching a cut-off is "grouping without resemblance", the opposite of the graded familial resemblance that characterizes prototypes (Wittgenstein 1953, Rosch 1978). There is an epistemic conflict between matching-to-prototype recognition, which is intrinsically graded, and a pass-or-fail diagnostic threshold (you are or are not autistic), which abolishes this gradation within a category. While prototype is based on family resemblance, the latter approach copies the necessary-sufficient framework that has proven to be appropriate only for mathematical fields, to work poorly in biology, and to have no psychological validity. It corresponds to a formal model of categorisation that fits neither with the way autism was discovered nor the psychological laws governing the use of concrete or abstract semantic entities when identifying a cluster of signs.

The application of prototype theory to an autism diagnosis grades the similarity of an individual to the subjective prototype of a limited number of experts who have long been exposed to an enriched population with suspected autism. We suggest replacing the reliance on increasing the N of studies incorporating heterogeneous individuals with increasing the N to whom the experts supervising recruitment have been exposed. What we lose in statistical power will be offset by a better signal-to-noise ratio, resulting from studying more resembling individuals.

Beyond the issue of gradation of familial resemblance, the notion of a prototype is associated with that of a basic level in a semantic hierarchy, in which the category maximises the information conveyed by correlated features. It coincides with the most frequently or precociously encountered set of features that discriminates one category from another at the same level. The mental image they evoke reflects the entire category without further analysis. The validity, the probability that a feature x predicts a category y and is not associated with another category, is maximized at the basic level. This notion of a basic level can be fruitfully

applied to psychiatric categories, which are organized hierarchically (Flanagan, Keeley, & Blashfield, 2012).