Tails or Types? A Critical Systematic Review of Taxometric Studies in Neurodevelopmental Disorders

Enrico Toffalini[1], Margherita Calderan[1], Riccardo Pagan[2], Valentina Tobia[3], and Edmund Sonuga-Barke[4]

[1]Department of General Psychology, University of Padua, Italy

[2]Department of Social and Developmental Psychology, University of Padua, Italy

[3]University Vita-Salute San Raffaele, Milan, Italy

[4]King's College London, UK

Author Note

Enrico Toffalini ⓘD https://orcid.org/0000-0002-1404-5133

Margherita Calderan ⓘD https://orcid.org/0009-0005-5668-5162

Correspondence concerning this article should be addressed to Enrico Toffalini, Email:

enrico.toffalini@unipd.it

Abstract

Background. Taxometric analysis is specifically designed to test whether the latent structures underlying clinical conditions are categorical (taxonic) or dimensional in nature. Despite its relevance to ongoing debates about the nature of neurodevelopmental conditions, its application to this field has been limited. **Methods.** We conducted a preregistered systematic review of published taxometric studies on autism, ADHD, learning disorders, and language impairment, with a focus on methodological threats such as artificial admixture (i.e., the practice of combining subsamples recruited via alternative methods, such as clinical and control groups, into a single dataset). A Monte Carlo simulation was run to illustrate how admixture can yield false taxonic results from dimensional data. **Results.** Fourteen articles met the inclusion criteria. Studies of ADHD and language impairment largely supported dimensional models, whereas most studies of autism and conditions' subtypes supported categorical models. However, admixture was common in studies reaching taxonic conclusions, especially for autism. In simulations, admixture was consistently associated with taxonic findings despite the true structure being dimensional. **Conclusions.** Evidence for categorical structure, especially in autism and disorders' subtypes, is likely undermined by methodological limitations, particularly admixture. We advocate for more high-quality taxometric studies in this area.

Keywords: Neurodevelopmental disorders, Taxometric analysis, Autism, ADHD, Learning disorders

Tails or Types? A Critical Systematic Review of Taxometric Studies in Neurodevelopmental Disorders

In recent years, there has been a reconsideration of the conceptual foundations of neurodevelopmental conditions (NDCs). The notion that diagnostic entities defined in the DSM-5 reflect discrete categories with identifiable non-arbitrary 'natural' boundaries,   separating disorder from non-disorder, and one disorder from another (American Psychiatric Association, 2013) has increasingly been questioned in favor of dimensional and transdiagnostic perspectives (e.g., Astle et al., 2022; Caviola et al., 2024; Happé & Frith, 2020, 2021; Mammarella et al., 2021; Michelini et al., 2024; Posner et al., 2020; Sonuga-Barke, 2020). In fact, the issue has been discussed for a long time. Sonuga-Barke (1998) questioned whether a medical model that frames childhood conditions as categorical disorders intrinsically linked to dysfunction was appropriate and useful. While diagnostic manuals such as the DSM are presented as atheoretical, they in practice adopt a "neo-Kraepelinian" vision of psychopathology, treating disorders as natural disease entities (see also Coghill & Sonuga-Barke, 2012).

A dimensional perspective reinterprets NDCs as extreme variations along continuous multivariate traits that are at least quasi-normally distributed in the general population. Accordingly, individuals with NDCs are not categorically distinct, and differ quantitatively rather than qualitatively from the neurotypical population. From this perspective diagnostic thresholds are essentially arbitrary, determined not by natural discontinuities but by pragmatic considerations such as clinical need (Arildskov et al., 2022; Haslam et al., 2006).

Determining whether NDCs represent tails of continuous distributions or discrete categories with boundaries grounded in biology or clinical reality, is important for our understanding of the ontological status of neurodevelopmental phenomena and for assessing the adequacy of the medical dysfunction paradigm. If NDCs are dimensional in nature, then the diagnostic thresholds used to categorize disorders are, by definition, arbitrary, reframing diagnostic categories as socially constructed groupings that include individuals at the margins of trait continua. The implications potentially affect how diagnosed individuals understand their

social identity, how clinicians make decisions, how classrooms and support systems are organized to accommodate neurodivergence, and how research is designed and interpreted (Caviola et al., 2024).

Support for a dimensional account of NDCs arises from several converging lines of evidence. First, individuals sharing the same diagnostic label often show substantial heterogeneity in symptom profiles, cognitive functioning, and developmental trajectories (e.g., Astle et al., 2022). Second, the boundaries between diagnostic categories are frequently blurred, with high comorbidity (Coghill & Sonuga-Barke, 2012; Michelini et al., 2024). Third, dimensional models are consistent with the growing evidence supporting multifactorial and polygenic etiology in NDCs (Astle & Fletcher-Watson, 2020; Caviola et al., 2024; Kovas & Plomin, 2006; Pennington et al., 2012; Plomin & Kovas, 2005). According to the central limit theorem (Fischer, 2011), when many individual causes (e.g., genetic variants) contribute additively, the resulting trait tends to approximate a continuous, Gaussian distribution. Despite this body of evidence, categorical thinking still dominates parts of the neurodevelopmental literature, for example concerning the categorical nature of ASD (e.g., Chown & Leatherland, 2021; Frazier et al., 2023; Happé & Frith, 2021), subtypes within NDCs (e.g., Burgess et al., 2018) or the proposed interpretation of emerging data-driven taxa (Astle et al., 2022).

While much of the evidence for dimensionality is circumstantial, taxometric analyses, originally introduced by Paul Meehl and colleagues (Meehl, 1995), were developed to provide a direct empirical test of whether latent clinical construct are categorical (taxonic) or dimensional in nature. Taxometric methods evaluate the structure of covariance among observed variables to determine whether there exists a latent boundary that divides individuals into qualitatively distinct groups. Conceptually, if such a boundary exists, it should alter and disrupt the patterns of relationships between indicators. Otherwise, the patterns are smooth and continuous across the entire range of observed values, supporting a dimensional model. Frequently used taxometric procedures include MAMBAC (Mean Above Minus Below A Cut), MAXCOV (Maximum Covariance), MAXEIG (Maximum Eigenvalue), and L-Mode (Latent Mode Factor Analysis)

(Cole, 2004; Ruscio, 2007; Ruscio & Ruscio, 2004b). Recently, simulation-based tools such as the Comparison Curve Fit Index (CCFI) have been introduced to provide objective, quantitative indices of fit between observed data and simulated taxonic versus dimensional comparison curves, facilitating inference (Ruscio et al., 2018).

Taxometric analyses have been widely applied in adult psychopathology, generally supporting dimensional structures for most adult psychiatric conditions (Haslam et al., 2012; Haslam et al., 2020). However, their use in neurodevelopmental research has remained limited, though the importance of taxometric methods as explicit tests for determining whether childhood disorders are categorical or dimensional has been repeatedly emphasized (Beauchaine, 2003; Coghill & Sonuga-Barke, 2012; Sonuga-Barke, 1998). In their debate on whether ASD reflects a category or a continuum, Happé and Frith (2021) and Chown and Leatherland (2021) do not mention such methods. In their review, Astle et al. (2022) briefly acknowledge taxometry, but predominantly emphasize studies using cluster analysis (a family of data-driven methods aimed at grouping individuals based on similarity across multiple variables). However, the limitations of clustering methods for assessing taxonicity have been discussed for decades: clustering algorithms tend to impose (rather than test for) categorical structures, they provide weak tests of taxonicity, and often fail to recover known taxa in simulations (Beauchaine, 2003; Ruscio & Ruscio, 2004b). Recent simulations using psychometrically realistic data suggest that commonly used clustering methods are prone to inflate the number of detected clusters (Toffalini et al., 2022, 2024). Indeed, Astle et al. (2022) interpreted clustering results as evidence that data-driven groups fail to align with DSM-5 categories, rather than as strong tests of taxonicity.

Like any method, taxometric analysis has its limitations. Commonly cited sources of bias in taxometric conclusions include poor indicator validity, nuisance covariance, low base rate, and artificial admixture (Cole, 2004; Haslam et al., 2020; Ruscio et al., 2018; Ruscio & Ruscio, 2004b, 2004a). Poor indicator validity implies that indicators present little mean difference between individuals with and without the target condition. Ideally, Cohen's *d* for each indicator should substantially exceed 1. Nuisance covariance refers to strong residual correlations among

indicators after accounting for group membership, indicating redundant information. Low base rate means that too few individuals in the sample belong to the taxon, limiting power to detect it (n > 30 for the taxonic group is usually considered necessary), which is problematic when studying rare conditions. Artificial admixture occurs when participants are combined from different recruitment sources or selection criteria, often as an attempt to compensate for a low base rate. Ruscio and Ruscio (2004a) describe three admixture-related sampling practices that may lead to pseudotaxonic results. First, when clinical and nonclinical participants are separately recruited and merged into a single sample. Second, when a sample is artificially split into subsamples based on different selection criteria, and these are then used to test different taxometric questions (cf. "subtractive" compound sampling, Haslam et al., 2020). Third, when researchers trim observations from the putative complement (individuals without a diagnosis) to inflate the base rate of the taxon. All these scenarios tend to artificially impose categorical structures onto data that might be strictly dimensional.

Goals of the Present Review

We aimed to critically review the use of taxometric methods in published studies of NDCs. We focused on the most commonly investigated categories: Autism Spectrum Disorder (ASD), Attention-Deficit/Hyperactivity Disorder (ADHD), Learning Disorders, and the closely related condition of Language Impairment/Disorder (the latter constitutes a minor deviation from the preregistration protocol). We excluded Intellectual Disability, whose etiology is often identifiable and taxonic/syndromic by definition (e.g., Down syndrome), or linked to known medical risk factors.

A crucial goal was to assess the methodological quality and credibility of taxometric findings. We specifically evaluated potential sources of bias, such as artificial admixture, and assessed whether studies provided sufficient methodological information to evaluate other known threats to validity, including low indicator validity, high nuisance covariance, and indicators skewness (extreme skewness can also create the illusion of categorical structures, Ruscio & Ruscio, 2004b; Ruscio & Ruscio, 2004a, and it might be relevant when assessing rare or extreme

behaviors).

Finally, we conduct a Monte Carlo simulation study to empirically illustrate the extent to which artificial admixture can lead to the risk to reaching taxonic conclusions in otherwise dimensional data. We simulate a dimensional scenario where clinical and nonclinical participants are separately recruited and merged into a single sample; taxometric (and model-based cluster) analyses are then performed on a set of observed indicators reflecting the latent clinical dimension.

Methods

The systematic review was conducted in accordance with the *Preferred Reporting items for Systematic Reviews and Meta-Analysis* (PRISMA) 2020 guidelines. A review protocol was preregistered on OSF registries at the following link:

https://osf.io/65y9g/?view_only=91e97c6b00dc444abb83fefcf67cbac0

Sources and Search strategy

A systematic search for relevant peer-reviewed articles was conducted in April 2025 using four electronic databases: PsycINFO, PubMed, Scopus, and Web of Science. No time limits were applied. The following search query, reflecting the methods and populations of interest, was applied to titles, abstracts, keywords:

```
(("taxometric*" OR "MAMBAC" OR "MAXCOV" OR "MAXEIG" OR "L-Mode") AND
("autis*" OR "ASD" OR "attention-deficit*" OR "ADHD" OR "learning dis*"
OR "reading dis*" OR "dyslexia" OR "math* dis*" OR "math* diff*"))
```

In PubMed, where the wildcard * was not supported, it was replaced with full terms (e.g., "disorder", "disability" instead of "dis*"; "mathematical", "mathematics", "math" instead of "math*").

Language disorder was subsequently added as a deviation from the preregistered protocol, replacing the second part of the search query with:

```
("language imp*" OR "language dis*")
```

Additionally, the full search query was launched on Google Scholar, and the first 10 pages of results were screened manually. References cited in, and articles citing, Haslam et al. (2020) were also screened via Google Scholar. Finally, an exploratory search using OpenAI's GPT-4o "Deep Research" tool was conducted, providing the preregistration protocol as context. All additional promising records identified through these methods were listed under the *"Identification of studies via other methods"* section of the flowchart.

Eligibility Criteria

Inclusion criteria were:

1. The study applied at least one taxometric procedure (MAMBAC, MAXCOV, MAXEIG, L-Mode, MAXSLOPE; other procedures would be considered if explicitly described as taxometric);
2. The target population consisted of individuals with NDCs/traits;
3. The study drew a conclusion about the latent structure (categorical vs. dimensional) or reported sufficient information (e.g., CCFIs) to infer a conclusion.

Exclusion criteria were:

1. The article was theoretical, methodological, simulation study without empirical data;
2. The study did not analyze behavioral data;
3. The study was a review and/or meta-analysis.

The search was limited to published, peer-reviewed journal articles, as we aimed to provide a critical review of the existing high-quality literature. Only one potentially relevant non-journal record was identified and, while excluded from the systematic review, is commented separately in the Discussion.

Screening and Coding

Two independent human coders (MC and RP) screened and coded all articles. All discrepancies were resolved through discussion and were later reviewed by the PI (ET).

Since a single article could report multiple (sub)samples and/or taxometric analyses, the full coded dataset followed a long-form data structure, with articles appearing across multiple rows. Coded variables were defined in the data dictionary included in the preregistration protocol, with a few additional fields added to capture coder notes and agreement metrics.

Inter-coder agreement was fair to good across all phases. In the title and abstract screening phase, Cohen's κ=0.79 (34 out of 38 independent decisions matched). In the full-text eligibility screening, Cohen's κ=0.52 (15 out of 19). For the final data extraction, 84% of rows (71 out of 85) were extracted by both coders. Agreement on critical variables was high: for the "authors' taxonic conclusion" agreement was 94%; for "artificial admixture", it was 95% at the (sub)sample level.

Results

A total of 110 records were identified through database searches. After removing 67 duplicates and excluding 5 additional records (dissertations or book chapters, only one of which appeared to include primary taxometric analyses but could not be retrieved), 38 articles were screened for title/abstract. Nineteen records were excluded at this stage. Full-text assessment was then conducted on the remaining 19 articles, of which 7 were excluded for not meeting the inclusion criteria. An additional 4 records were identified through citation tracking and online browsing, of which 2 were included in the final review. The PRISMA flowchart in Figure 1 summarizes the study selection process.

Figure 1 *about here*

A summary of the included studies, simplified to one row per article, is presented in Table 1. A full reporting of all coded data for each individual taxometric analysis is available in the Supplemental Online Material. The total number of taxometric analyses was 85, ranging from 1 to 22 per article. Targeted disorders were ADHD (6 articles), ASD (6 articles), dyslexia (1 article), and Specific Language Impairment (SLI; 2 articles), with one article addressing both ADHD and ASD. Most articles (10 out of 14) examined whether a given disorder constituted a taxon, while the remaining 4 focused on whether putative subtypes within a disorder showed taxonic structure.

The most commonly used taxometric method was MAMBAC, applied in 13 out of 14

articles and in 84 out of 85 total analyses. This was followed by MAXEIG (9 articles), L-Mode (7 articles), MAXCOV (2 articles), and MAXSLOPE (1 article). Most studies employed more than one taxometric procedure, with a median of 3 methods per analysis. In 11 out of 14 articles, interpretation was based on a quantitative fit index (Fit$_d$ or, in more recent studies, the CCFI), while the remaining 3 relied on visual inspection of taxometric curves. The number of indicators entered into analyses ranged from 2 to 9, with a median of 3 (excluding one article in which the number was not clearly reported). The median sample size across all analyses was 600, or 611.25 when averaged within articles.

Taxonic conclusions were mixed. All five articles targeting ADHD as a disorder concluded in favor of a dimensional structure, as did both articles focusing on SLI. In contrast, all four articles examining ASD as a disorder leaned toward, or explicitly supported, a categorical (taxonic) structure. The remaining four articles investigated subtypes of disorders, and all favored taxonic conclusions, either predominantly (Stevens et al., 2018 on ADHD subtypes; Munson et al., 2008 on ASD subtypes; O'Brien et al., 2012 on dyslexia subtypes), or partially (Ingram et al., 2008 supported taxonic ASD subtypes depending on the set of indicators used). In total, 8 out of 14 articles reached a taxonic conclusion in at least some of their taxometric analyses. Specifically, 28 out of 85 total analyses—representing 32.9%—yielded taxonic results. These were primarily associated with ASD: 21 out of 29 analyses involving ASD (i.e., 72.4%) supported a categorical structure.

Artificial admixture was evident or strongly suspected in at least part of the analyses in 8 out of 14 articles (including all those targeting ASD) and involving a total of 31 out of 85 analyses. Articles that incurred artificial admixture tended to align with those that reached taxonic conclusions: 6 of the 8 articles with suspected admixture also reported taxonic findings. This association was even more pronounced at the level of individual analyses. A taxonic conclusion was reached in 20 out of 31 analyses (64.5%) where admixture was evident or suspected. In contrast, only 8 out of 54 analyses (14.8%) reached a taxonic conclusion when admixture was likely or certainly

Other methodological issues concerned missing information about key descriptive statistics and validity indices. Nuisance covariance could be coded for only 2 out of 14 articles. Where reported, it was small to negligible in James et al. (2016) , but exceeded 0.40 for many indicators in Frazier et al. (2023) . Both studies reached a taxonic conclusion. Skewness of indicator score distributions was reported in only 5 articles, where skewness values were often near or beyond ±1 for at least one indicator in about one-third of analyses. However, there was no clear association between skewness and whether a taxonic conclusion was reached. By contrast, indicator validity was more consistently reported, and could be coded in all but 3 articles. Where available, it was almost always large or very large (Cohen's d 1), consistent with best-practice recommendations for taxometric analysis.

Table 1 *about here*

Additional Use of Cluster Analysis

Among the 14 identified articles, 6 also employed clustering methods: in 5 cases consisting of latent class/profile analysis (LCA, LPA); the remaining case employing community detection in a graph analysis. In 5 of these 6 articles clustering was conducted in parallel with taxometric analysis (Deserno et al., 2023; Frazier et al., 2007; Frazier et al., 2023; James et al., 2016; Munson et al., 2008), while in one case it represented a secondary analysis (Frazier et al., 2010). All six studies targeted ADHD and/or ASD, and in all cases clustering results were interpreted as supporting taxonic conclusions. These interpretations generally converged with those from taxometric analysis, with one exception: in Frazier et al. (2007), clustering results inconclusively suggested the presence of subgroups within ADHD, contrasting with the dimensional structure favored by the taxometric analysis (it is worth reminding that clustering methods do not provide explicit tests of taxonicity). In all but two studies (Frazier et al., 2007; Munson et al., 2008), we identified at least some risk of artificial admixture.

A Monte Carlo Simulation of Artificial Admixture

Artificial admixture represents a common and powerful threat to validity, but also one that is easily avoidable by design. In our review, it presented a marked risk of bias for studies

supporting the taxonic nature of autism. To illustrate the risks associated with artificial admixture, we simulate a dimensional scenario where clinical and nonclinical participants are separately recruited and merged into a single sample

A single normally distributed latent variable X is generated to represent the clinical trait of interest. A set of observed indicators (x_1, x_2, x_3) are then simulated as linear functions of X plus normally distributed random noise, yielding $R^2 = 0.77$ for each indicator. Individuals above the 99th percentile of X are assumed to receive a diagnosis via global clinical assessment. A sample of $N = 1000$ individuals is drawn from the general population. Given the low base rate, only about $n = 10$ cases are expected to naturally receive a diagnosis. To address this, researchers draw a second sample consisting of an additional $n = 1000$, targeting clinically diagnosed cases, and merge it with the original sample. The full simulation code and data are available in the online materials:

https://osf.io/ys5ad/?view_only=0e2812acfabf43fc8af01f883c7331ba

An instance of the simulation, illustrating two observed variables x_1 and x_2, is presented in Figure 2. Color is used to distinguish individuals with and without a diagnosis, but even without this aid, two clusters visually emerge. This apparent separation is misleading: no latent categories exist in the data. The appearance of a categorical structure is a consequence of artificial admixture. In this simulated example, indicator validity is high (Cohen's $d \approx 2$), suggesting strong separation between groups on each variable. However, nuisance covariance is partly problematic: among diagnosed individuals, the correlation between indicators is close to zero ($r \approx 0.05$), whereas among non-diagnosed individuals, it is substantial ($r \approx 0.60$).

Figure 2 *about here*

A Monte Carlo simulation was then conducted with 1000 iterations. Taxometric analyses were performed using the `RTaxometric` package in R. Inference was based on the Comparison Curve Fit Index (CCFI), which ranges from 0 (indicating strong support for a dimensional structure) to 1 (strong support for a categorical structure), with values around 0.5 considered inconclusive. The default procedures (MAMBAC, MAXEIG, and L-Mode) were applied to each

simulated dataset. To enable all three procedures, three observed indicators (x_1, x_2, x_3) were included in the simulations.

Over 1000 simulated datasets, results consistently and strongly favored a categorical interpretation: median combined CCFI = 0.88, 95% quantile interval [0.83, 0.92]; median MAMBAC CCFI = 0.90; average MAXEIG CCFI= 0.86; average L-Mode CCFI = 0.88. The smallest combined CCFI over 1000 iterations was 0.80. Therefore, admixture can be an extremely strong driver of false categorical findings, especially when many clinical cases are separately recruited and added to a sample.

To verify that artificial admixture also produces misleading results when using clustering methods, we ran Gaussian Mixture Models (GMMs) comparing 1-component versus 2-component solutions, using the `mclust` package in R. Model selection was based on the Bayesian Information Criterion (BIC), as per default in GMMs. GMM is a form of model-based clustering grounded in finite mixture modeling. It is conceptually similar to latent profile analysis (LPA; with which it is often confused) but is more flexible in that it models covariances among indicators. Across 1,000 simulated datasets, results consistently favored the 2-component solution over the unimodal alternative. The median difference in BIC (ΔBIC) was -548.67, in favor of the 2-component model.

To confirm that the categorical structure was incorrectly favored specifically due to artificial admixture, we run a second Monte Carlo simulation using the same dimensional data-generating process as before, but without admixture. In this case, a single sample of N = 2000 observations was drawn at each iteration. Therefore, nuisance covariance remains present, but there is no admixture. Median combined CCFI = 0.47, 95% quantile interval [0.40, 0.58]; median MAMBAC CCFI = 0.39; average MAXEIG CCFI= 0.49; average L-Mode CCFI = 0.53. Clustering with GMM also consistently favored the 1-component model (over the 2-component model), with an average ΔBIC = -30.36 in favor of the 1-component model.

Thus, without admixture taxometric results leaned towards a dimensional conclusion, but frequently remained inconclusive, probably due to nuisance covariance. It is worth noting,

however, that under a dimensional structure achieving both high indicator validity and low within-group covariance might be practically impossible, because indicators that strongly differentiate individuals with and without a diagnosis must be strongly associated with the underlying trait, thus producing inter-indicator covariance.

Discussion

We found that taxometric analysis remains underused in neurodevelopmental research, which is unfortunate given the unique relevance of this method to the ongoing debate about the nature of NDCs (Astle et al., 2022; Happé & Frith, 2021; Michelini et al., 2024; Sonuga-Barke, 1998, 2020). That underuse was expected, however. Haslam et al. (2020) classified only 15 out of 183 taxometric studies in the "childhood" category (which included heterogeneous conditions, and even sleep problems). In the studies identified in our review, only ASD, ADHD, and SLI were targeted as disorders in taxometric analysis. None of learning disorders employed taxometric approaches, with Dyslexia being covered in one single article, but only for investigating its subtypes.

Taxometric results appeared relatively consistent. Dimensional conclusions were predominantly reached for ADHD and SLI, while categorical conclusions prevailed for ASD and studies examining subtypes of any disorder. This may seem to offer a compelling answer to the debate raised by Chown and Leatherland (2021) and Happé and Frith (2021) However, a closer inspection of the methodological limitations in the reviewed studies suggests that such conclusions should be interpreted with caution.

We identified four ASD taxometric analyses, and all of them reached taxonic conclusions, with some degree of artificial admixture. This represents a major threat to validity of the conclusions of these studies, as demonstrated both in our data simulation and in prior literature (Ruscio & Ruscio, 2004b, 2004a). Admixture in the reviewed studies ranged from sampling participants through different methods, only some of which explicitly targeted clinical populations (James et al., 2016), to combining clinical and non-clinical samples (Deserno et al., 2023; Frazier et al., 2023). In one case, inclusion criteria from a larger database were partly unclear, but the

sample was explicitly structured by diagnostic status (Deserno et al., 2023). Another form of admixture involved constructing a sample of affected and unaffected siblings (Frazier et al., 2010). According to our interpretation, the only relatively unbiased taxometric evidence suggesting that ASD may be a taxon as a disorder comes from a secondary analysis by Frazier et al. (2010), in which only unaffected siblings were examined.

Concerning subtypes, the situation was similar. All four reviewed articles reached some taxonic conclusions, but three of them likely involved some form of artificial admixture. However, the issue appeared to be nuanced and less severe in this subset of studies. For example, O'Brien et al. (2012) used psychometric cutoffs across different tools and alternate inclusion criteria, which may not necessarily have biased the results, unless the selection criteria correlate with different underlying cognitive impairments in dyslexia (phonological vs. non-phonological). However, the authors themselves acknowledged this possibility, writing that *"depending on the dyslexia criterion used, samples may vary in composition with regard to their subtypes"* (p.33). A subtle and potentially overlooked cause of admixture in studies targeting subtypes within entirely clinical samples is that the original diagnoses might be based on exceeding cutoff thresholds on partly independent traits or tasks, reflecting a variant of the second type of problematic sampling listed by Ruscio and Ruscio (2004a) In general, a taxonic structure of subtypes should be regarded as inconsistent, and in principle unlikely, if the main condition as a whole is seen as dimen

Overall, the presence of artificial admixture was most frequently associated with taxonic conclusions. Only two notable mismatches were identified: Marcus et al. (2012) reached a dimensional conclusion regarding ADHD despite having incurred admixture of clinical cases, while Munson et al. (2008) reached a taxonic conclusion about ASD subtypes without evidence of admixture (all children had received an ASD diagnosis based on a seemingly uniform clinical assessment). Together with the above discussed exception in a secondary analysis of a subsample by Frazier et al. (2010) targeting ASD, these findings reinforce the view that ADHD is best conceptualized as dimensional, while ASD may exhibit certain taxonic features. However, the latter conclusion is supported by limited unbiased evidence.

In our simulated example, we showed that artificial admixture represents a major threat to the validity of taxometric analysis, and to clustering methods as well. One might argue that our simulated scenario, in which half of sample was separately recruited, was extreme. However, this design is broadly consistent with the degree of admixture observed in the reviewed literature.

Conclusions, Limitations, Future Directions

An obvious limitation of the present review concerns the small number of studies identified, and the virtual lack of studies exploring taxa within learning disorders. Furthermore, the available evidence was too scattered and methodologically heterogeneous to conduct a meta-analysis. These limitations are themselves informative, highlighting the need for more taxometric investigation in the field. It is unlikely that many more studies would be included by relaxing eligibility criteria. Exclusion criteria specified the lack of use of behavioral indicators; however, no study was excluded on this basis. One record was screened out for being a dissertation (Clemons, 2006); however, the full document could not be retrieved. Based on its abstract, it focused on whether ADHD subtypes represent distinct taxa, and it reached a taxonic conclusion, but some degree of admixture was possible (as participants were separately recruited from clinical centers and schools).

A relevant future direction concerns a systematic examination of the use and results of cluster analysis, which was employed by some of the reviewed articles, and is a prominent tool in this field (Astle et al., 2022). Model-based approaches, such as latent class/profile analysis, were the most frequently used. Clustering algorithms are not without problems. First, while they are more flexible and can address broader questions, they do not offer direct or consistent tests of taxonicity (Beauchaine, 2003; Ruscio & Ruscio, 2004b). Second, meaningful scientific inference from clustering (beyond its use as a mere data reduction technique) requires careful handling of assumptions, as the risk of detecting pseudoclusters is high (Toffalini et al., 2022, 2024). Finally, it is not immune to risks such as artificial admixture. Nonetheless, clustering and taxometric analysis could be more directly, and critically, compared when addressing the question of whether neurodevelopmental conditions reflect categories or dimensions.

A crucial implication of our work concerns how taxometric findings should inform our interpretation of traditional diagnostic categories of NDCs such as those in the DSM-5, which have come under criticism (Astle et al., 2022; Sonuga-Barke, 2020, 2022). Although described as atheoretical and agnostic about etiology, the DSM operationalizes disorders in ways that reinforce a neo-Kraepelinian medical paradigm, implicitly reflecting the idea that NDCs reflect discrete categories underpinned by well-defined, specific dysfunctions. This view continues to shape research and clinical practice (Coghill & Sonuga-Barke, 2012). Finding that the latent structure of one or many NDCs is dimensional does not deny the practical utility of diagnostic systems. However, it requires researchers and clinicians to be explicit about their assumptions and models, instead of implicitly reifying categories as natural taxa and overinterpreting their ontological status. High-quality taxometric studies offer a direct solution of the issue. We argue that they deserve a more central role in the study of NDCs.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5.* (5th edition). American Psychiatric Association.

Arildskov, T. W., Sonuga-Barke, E. J. S., Thomsen, P. H., Virring, A., & Østergaard, S. D. (2022). How much impairment is required for ADHD? No evidence of a discrete threshold. *Journal of Child Psychology and Psychiatry*, *63*(2), 229–237. https://doi.org/10.1111/jcpp.13440

Astle, D. E., & Fletcher-Watson, S. (2020). Beyond the Core-Deficit Hypothesis in Developmental Disorders. *Current Directions in Psychological Science*, *29*(5), 431–437. https://doi.org/10.1177/0963721420925518

Astle, D. E., Holmes, J., Kievit, R., & Gathercole, S. E. (2022). Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *63*(4), 397–417. https://doi.org/10.1111/jcpp.13481

Beauchaine, T. P. (2003). Taxometrics and developmental psychopathology. *Development and Psychopathology*, *15*(3), 501–527. https://doi.org/10.1017/s0954579403000270

Burgess, A. P., Witton, C., Shapiro, L., & Talcott, J. B. (2018). *From Subtypes to Taxons:*

Identifying Distinctive Profiles of Reading Development in Children (T. Lachmann & T. Weis, Eds.; pp. 213–233). Springer International Publishing. https://doi.org/10.1007/978-3-319-90805-2_11

Caviola, S., Greiff, S., & Toffalini, E. (2024). Learning disorders and difficulties: From a categorical to a dimensional perspective. *Learning and Individual Differences*, *113*, 102490. https://doi.org/10.1016/j.lindif.2024.102490

Chown, N., & Leatherland, J. (2021). Can a Person be 'A Bit Autistic'? A Response to Francesca Happé and Uta Frith. *Journal of Autism and Developmental Disorders*, *51*(2), 749–751. https://doi.org/10.1007/s10803-020-04541-0

Clemons, S. D. (2006). *A taxometric comparison of attention deficit hyperactive disorder: Inattentive subtype and combined subtype.* [PhD thesis]. https://research.ebsco.com/linkprocessor/plink?id=03d321c9-6249-390d-9639-9de020c77f90

Coghill, D., & Sonuga-Barke, E. J. S. (2012). Annual Research Review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders – implications of recent empirical study. *Journal of Child Psychology and Psychiatry*, *53*(5), 469–489. https://doi.org/10.1111/j.1469-7610.2011.02511.x

Cole, D. A. (2004). Taxometrics in psychopathology research: an introduction to some of the procedures and related methodological issues. *Journal of Abnormal Psychology*, *113*(1), 3–9. https://doi.org/10.1037/0021-843X.113.1.3

Deserno, M. K., Bathelt, J., Groenman, A. P., & Geurts, H. M. (2023). Probing the overarching continuum theory: data-driven phenotypic clustering of children with ASD or ADHD. *European Child & Adolescent Psychiatry*, *32*(10), 1909–1923. https://doi.org/10.1007/s00787-022-01986-9

Fischer, H. (2011). *A History of the Central Limit Theorem: From Classical to Modern Probability Theory*. Springer. https://doi.org/10.1007/978-0-387-87857-7

Frazier, T. W., Chetcuti, L., Al-Shaban, F. A., Haslam, N., Ghazal, I., Klingemier, E. W., Aldosari, M., Whitehouse, A. J. O., Youngstrom, E. A., Hardan, A. Y., & Uljarević, M. (2023).

Categorical versus dimensional structure of autism spectrum disorder: A multi-method investigation. *JCPP Advances*, *3*(2), e12142. https://doi.org/10.1002/jcv2.12142

Frazier, T. W., Youngstrom, E. A., & Naugle, R. I. (2007). The latent structure of attention-deficit/hyperactivity disorder in a clinic-referred sample. *Neuropsychology*, *21*(1), 45–64. https://doi.org/10.1037/0894-4105.21.1.45

Frazier, T. W., Youngstrom, E. A., Sinclair, L., Kubu, C. S., Law, P., Rezai, A., Constantino, J. N., & Eng, C. (2010). Autism Spectrum Disorders as a Qualitatively Distinct Category From Typical Behavior in a Large, Clinically Ascertained Sample. *Assessment*, *17*(3), 308–320. https://doi.org/10.1177/1073191109356534

Happé, F., & Frith, U. (2020). Annual Research Review: Looking back to look forward – changes in the concept of autism and implications for future research. *Journal of Child Psychology and Psychiatry*, *61*(3), 218–232. https://doi.org/10.1111/jcpp.13176

Happé, F., & Frith, U. (2021). Dimensional or Categorical Approaches to Autism? Both are Needed. A Reply to Nick Chown and Julia Leatherland. *Journal of Autism and Developmental Disorders*, *51*(2), 752–753. https://doi.org/10.1007/s10803-020-04728-5

Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: a quantitative review of taxometric research. *Psychological Medicine*, *42*(5), 903–920. https://doi.org/10.1017/S0033291711001966

Haslam, N., McGrath, M. J., Viechtbauer, W., & Kuppens, P. (2020). Dimensions over categories: a meta-analysis of taxometric research. *Psychological Medicine*, *50*(9), 1418–1432. https://doi.org/10.1017/S003329172000183X

Haslam, N., Williams, B., Prior, M., Haslam, R., Graetz, B., & Sawyer, M. (2006). The Latent Structure of Attention-Deficit/Hyperactivity Disorder: A Taxometric Analysis. *Australian & New Zealand Journal of Psychiatry*, *40*(8), 639–647. https://doi.org/10.1080/j.1440-1614.2006.01863.x

Ingram, D. G., Takahashi, T. N., & Miles, J. H. (2008). Defining autism subgroups: a taxometric solution. *Journal of Autism and Developmental Disorders*, *38*(5), 950–960.

https://doi.org/10.1007/s10803-007-0469-y

James, R. J. E., Dubey, I., Smith, D., Ropar, D., & Tunney, R. J. (2016). The Latent Structure of
Autistic Traits: A Taxometric, Latent Class and Latent Profile Analysis of the Adult Autism
Spectrum Quotient. *Journal of Autism and Developmental Disorders*, *46*(12), 3712–3728.
https://doi.org/10.1007/s10803-016-2897-z

Kovas, Y., & Plomin, R. (2006). Generalist genes: implications for the cognitive sciences. *Trends
in Cognitive Sciences*, *10*(5), 198–203. https://doi.org/10.1016/j.tics.2006.03.001

Mammarella, I. C., Toffalini, E., Caviola, S., Colling, L., & Szűcs, D. (2021). No evidence for a
core deficit in developmental dyscalculia or mathematical learning disabilities. *Journal of
Child Psychology and Psychiatry*, *62*(6), 704–714. https://doi.org/10.1111/jcpp.13397

Marcus, D. K., Norris, A. L., & Coccaro, E. F. (2012). The latent structure of attention
deficit/hyperactivity disorder in an adult sample. *Journal of Psychiatric Research*, *46*(6),
782789.

Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in
psychopathology. *American Psychologist*, *50*(4), 266–275.

Michelini, G., Carlisi, C. O., Eaton, N. R., Elison, J. T., Haltigan, J. D., Kotov, R., Krueger, R. F.,
Latzman, R. D., Li, J. J., Levin-Aspenson, H. F., Salum, G. A., South, S. C., Stanton, K.,
Waldman, I. D., & Wilson, S. (2024). Where do neurodevelopmental conditions fit in
transdiagnostic psychiatric frameworks? Incorporating a new neurodevelopmental spectrum.
World Psychiatry, *23*(3), 333–357. https://doi.org/10.1002/wps.21225

Munson, J., Dawson, G., Sterling, L., Beauchaine, T., Zhou, A., Koehler, E., Lord, C., Rogers, S.,
Sigman, M., Estes, A., & Abbott, R. (2008). Evidence for latent classes of IQ in young
children with autism spectrum disorder. *American Journal on Mental Retardation*, *113*(6),
439–452. https://doi.org/10.1352/2008.113:439-452

O'Brien, B. A., Wolf, M., & Lovett, M. W. (2012). A Taxometric Investigation of Developmental
Dyslexia Subtypes. *Dyslexia*, *18*(1), 16–39. https://doi.org/10.1002/dys.1431

Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A.,

Leopold, D., Samuelsson, S., Byrne, B., Willcutt, E. G., & Olson, R. K. (2012). Individual prediction of dyslexia by single vs. Multiple deficit models. *Journal of Abnormal Psychology*, *121*(1), 212–224. https://doi.org/10.1037/a0025823

Plomin, R., & Kovas, Y. (2005). Generalist Genes and Learning Disabilities. *Psychological Bulletin*, *131*(4), 592–617. https://doi.org/10.1037/0033-2909.131.4.592

Posner, J., Polanczyk, G. V., & Sonuga-Barke, E. (2020). Attention-deficit hyperactivity disorder. *Lancet*, *395*(10222), 450–462. https://doi.org/10.1016/S0140-6736(19)33004-1

Ruscio, J. (2007). Taxometric Analysis: An Empirically Grounded Approach to Implementing the Method. *Criminal Justice and Behavior*, *34*(12), 1588–1622. https://doi.org/10.1177/0093854807307027

Ruscio, J., Carney, L. M., Dever, L., Pliskin, M., & Wang, S. B. (2018). Using the comparison curve fix index (CCFI) in taxometric analyses: Averaging curves, standard errors, and CCFI profiles. *Psychological Assessment*, *30*(6), 744–754. https://doi.org/10.1037/pas0000522

Ruscio, J., & Ruscio, A. M. (2004a). A conceptual and methodological checklist for conducting a taxometric investigation. *Behavior Therapy*, *35*(2), 403–447. https://doi.org/10.1016/S0005-7894(04)80044-3

Ruscio, J., & Ruscio, A. M. (2004b). Clarifying Boundary Issues in Psychopathology: The Role of Taxometrics in a Comprehensive Program of Structural Research. *Journal of Abnormal Psychology*, *113*(1), 24–38. https://doi.org/10.1037/0021-843x.113.1.24

Sonuga-Barke, E. J. S. (1998). Categorical Models of Childhood Disorder: A Conceptual and Empirical Analysis. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(1), 115–133. https://doi.org/10.1017/S0021963097001728

Sonuga-Barke, E. J. S. (2020). Editorial: 'People get ready': Are mental disorder diagnostics ripe for a Kuhnian revolution? *Journal of Child Psychology and Psychiatry*, *61*(1), 1–3. https://doi.org/10.1111/jcpp.13181

Sonuga-Barke, E. J. S. (2022). 'If it don't fit, don't force it?' If real-world, complex clinical decisions are intrinsically categorical can dimensional systems add value? Reflections on

lahey, tiemeier & krueger (2022). *JCPP Advances*, *2*(4), e12120.

https://doi.org/10.1002/jcv2.12120

Stevens, M. C., Pearlson, G. D., Calhoun, V. D., & Bessette, K. L. (2018). Functional

neuroimaging evidence for distinct neurobiological pathways in attention-deficit/hyperactivity

disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(8), 675–685.

https://doi.org/10.1016/j.bpsc.2017.09.005

Toffalini, E., Gambarota, F., Perugini, A., Girardi, P., Tobia, V., Altoè, G., Giofrè, D., Team, P. C.,

& Feraco, T. (2024). Clusters that are not there: An R tutorial and a Shiny app to quantify a

priori inferential risks when using clustering methods. *International Journal of Psychology*,

59(6), 1183–1198. https://doi.org/10.1002/ijop.13246

Toffalini, E., Girardi, P., Giofrè, D., & Altoè, G. (2022). Entia Non Sunt Multiplicanda … Shall I

look for clusters in my cognitive data? *PLOS ONE*, *17*(6), e0269584.

https://doi.org/10.1371/journal.pone.0269584

Table 1

Overview summary of the reviewed articles.

Study	Target disorder	Taxometric focus	Population analyzed	Age	Taxonic conclusion	Artificial admixture	Sample size	No. of taxometric analyses	Taxometric methods	Interpret. based on	No. of indicators	Indicators skewn.	Nuisance covar.	Indicators validity
Marcus et al. (2012)	ADHD	disorder	Mixed clinical and non-clinical	adults	no	yes	432 to 600	6	MAMBAC; MAXEIG; L-Mode	CCFI	2 to 9	unreported	unreported	2.04 to 3.26
Haslam et al. (2006)	ADHD	disorder	General population	children	no	no	1222 to 1774	2	MAXEIG; MAMBAC	Fitd	3	0.45 to 1.49	unreported	2.47 to 3.29
Frazier et al. (2007)	ADHD	disorder	Clinical referrals	children to adults	no	no	191 to 394	22	MAMBAC; MAXEIG; L-Mode	Fitd	3 to 9	-1.09 to 0.47	unreported	1.03 to 3.59
Marcus and Barry (2011)	ADHD	disorder	General population	children	no	no	667 to 1078	15	MAMBAC; MAXEIG; L-Mode	CCFI	2 to 9	0.74 to 3.91	unreported	1.96 to 3.26
Stevens et al. (2018)	ADHD	subtypes	Mixed clinical and non-clinical	adolescents	yes	partly yes	71 to 251	2	MAMBAC; MAXS-LOPE	CCFI	3	unreported	unreported	unreported
Deserno et al. (2023)	ADHD / ASD	disorder	Mixed clinical and non-clinical	children	no for ADHD, mostly yes for ASD	likely yes	203 to 434	4	L-Mode; MAMBAC; MAXEIG	CCFI	3	unreported	unreported	1.86 to 2.74
Frazier et al. (2023)	ASD	disorder	Mixed clinical and non-clinical	children to adults	yes	yes	512 to 16755	10	MAMBAC; MAXEIG; L-Mode	CCFI	3 to 8	0.19 to 0.99	0.14 to 0.73	0.57 to 1.84
Frazier et al. (2010)	ASD	disorder	Clinical and unaffected siblings	children	yes	partly yes	1825 to 11472	6	MAMBAC; MAXEIG	CCFI	3 to 4	unreported	unreported	unreported
James et al. (2016)	ASD	disorder	General population	adults	yes	yes	1139	1	MAMBAC; MAXCOV; MAXEIG; L-Mode	CCFI	8	-0.48 to 0.73	-0.09 to 0.28	0.15 to 2.21
Munson et al. (2008)	ASD	subtypes	Clinical	children	yes	no	456	1	MAXCOV	visual	4	unreported	unreported	unreported
Ingram et al. (2008)	ASD	subtypes	Clinical	children	partly yes	no	481 to 2254	7	MAMBAC; MAXCOV	CCFI	unclear	unreported	unreported	1.51 to 2.55
O'Brien et al. (2012)	Dyslexia	subtypes	Clinical	children	mostly yes	partly yes	517 to 671	6	MAMBAC; MAXEIG; L-Mode	CCFI	3 to 4	unreported	unreported	0.82 to 2.37
Dollaghan (2004)	SLI	disorder	General population	children	no	no	620 to 623	2	MAMBAC	visual	2	unreported	unreported	1.93 to 2.81
Dollaghan (2011)	SLI	disorder	General population	children	no	no	601	1	MAMBAC	visual	2	unreported	unreported	3.47 to 3.52

Figure 1

PRISMA 2020 flow diagram illustrating the study screening and selection process
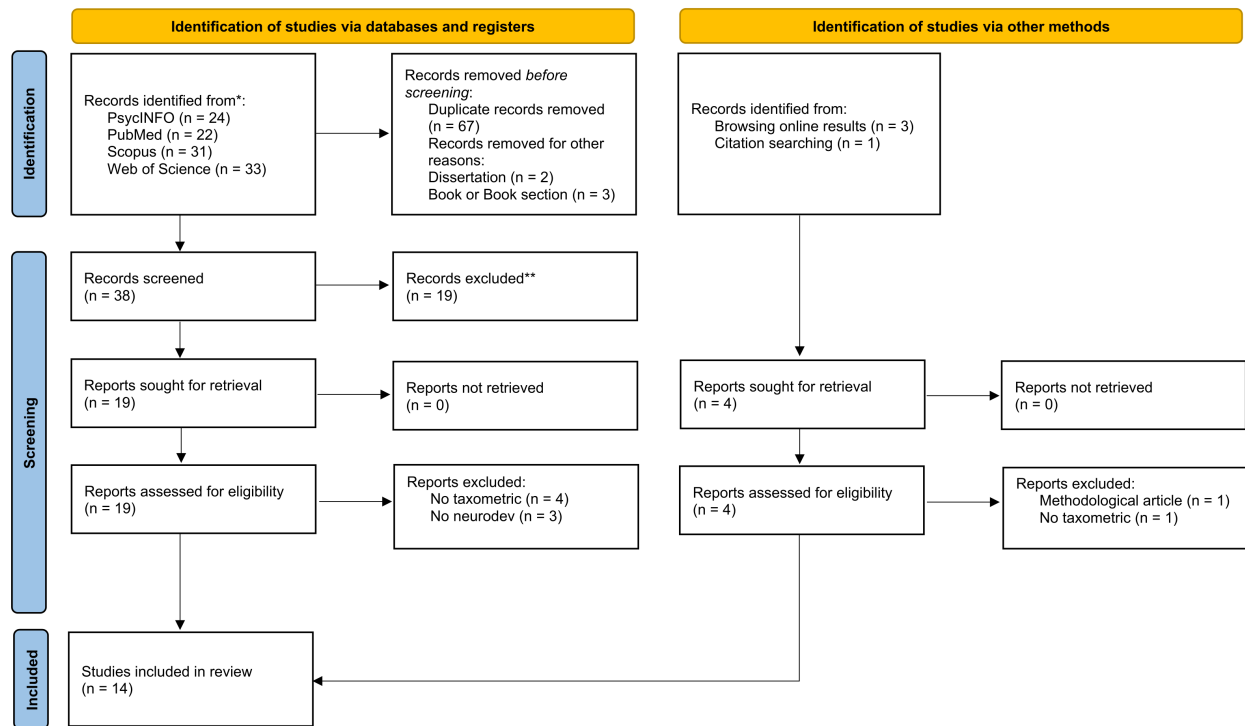
Figure 2

A simulated example of artificial admixture, where a large number of individuals with a diagnosis are separately recruited and added to the sample (total N = 2000). The true underlying structure is strictly dimensional.