TPM®

GAUSSIAN MIXTURE MODELS FOR THE ANALYSIS OF WISC-IV DIMENSIONS: A MULTIVARIATE APPROACH TO IMPROVE THE ASSESSMENT OF INTELLECTUAL FUNCTIONING

ROSA FABBRICATORE
DAVIDE MAROCCO
FRANCESCO PALUMBO
UNIVERSITY OF NAPOLI FEDERICO II

SERAFINO BUONO
I.R.C.C.S. OASI MARA SS., TROINA

SANTO DI NUOVO
UNIVERSITY OF CATANIA

The Wechsler Intelligence Scale for Children-IV provides four indexes that analyze the intellectual functioning in specific cognitive fields and a full-scale intelligence quotient (FSIQ) as measure of the general cognitive ability. However, often the diagnostic process considers the FSIQ score only. This study exploits the Gaussian mixture model (GMM) as a statistical tool to analyze WISC-IV capability to support the diagnostic decision-making process in a multidimensional approach based on the joint evaluation of the four main indexes. The study was conducted on two groups of participants (10 and 12 years old with N = 52 and N = 47, respectively) with clinical diagnosis. In addition, N = 50 observations were randomly generated from the distribution of the Italian reference populations referred to each age group. In both groups, GMM detected two components underlining different behaviors in central tendency, variability, and correlation. Comparison of GMM partitions with a supervised classification shows that group memberships are congruent.

Keywords: Intellectual functioning; WISC-IV; Multivariate assessment; Gaussian mixture models; Intellectual disabilities.

Correspondence concerning this article should be addressed to Rosa Fabbricatore, Department of Social Sciences, University of Napoli Federico II, Vico Monte della Pietà 1, 80138 Napoli (NA), Italy. Email: rosa.fabbricatore@unina.it
Most of the work done by Dr. Rosa Fabbricatore was carried out while she was scholarship holder at the Department of Political Sciences, University of Napoli Federico II.

Wechsler scales are currently recognized as one of the most reliable psycho-diagnostic tools to describe and measure intellectual functioning (Evers et al., 2012). The Wechsler Intelligence Scale for Children-IV (WISC-IV) has been assessed to measure the intellectual ability of children between 6 years and 0 months and 16 years and 11 months old (Wechsler, 2003). The scale consists of 10 core subtests and five additional ones that can be used to obtain a more detailed description of cognitive functioning or to substitute comparable core subtest, when the latter is not available, or for test-retest purposes.

Like most of the psycho-diagnostic scales, the WISC-IV grounds on the assumption that intelligence is a multidimensional construct and can be defined as the integration of four scores. In detail, the scale provides both an overall measure of the general cognitive ability (the full scale IQ) and four indexes that analyze the intellectual functioning in specific cognitive fields that are: verbal comprehension (VC), perceptual reasoning (PR), working memory (WM), and processing speed (PS).

The primary use of WISC-IV is diagnostic assessment (Lang, Di Pierro, Michelotti, & Squarza, 2017): the scale allows clinicians to identify subjects who are lagging in cognitive development or who have a decline in their intellectual functioning. However, often during the diagnostic process, especially for the evaluation of intellectual disability for assigning specialized support in the school, only the final full-scale score is considered to diagnose the global intellectual disabilities, ignoring the four subscale partial scores.

Moreover, another worrying aspect regarding the WISC-IV use in the diagnostic field is that the normative parameters used for the diagnosis only refer to IQ scores registered on a population recognized as "normal intellectual functioning." Then, according to the normative parameters, subjects with intellectual disabilities are all located under the threshold of μ–2σ (i.e., full-scale intelligent quotient, FSIQ = 70). For this reason, to our knowledge, the specialized literature does not provide guidelines to recognize and correctly classify the profiles of subjects with intellectual disabilities that consequently result strongly flattened. Indeed, IQ measures are less valid in the lower end of the IQ range (American Psychiatric Association, 2013).

It is worth remembering that before the publication of the fifth revision of the diagnostic and statistical manual of mental disorders (DSM-5), intellectual disabilities were diagnosed according to the WISC-IV full scale IQ, where the threshold was set to two standard deviations below the mean (American Psychiatric Association, 2000).

After the publication of DSM -5 (American Psychiatric Association, 2013), deficits in intellectual functions are a necessary but not sufficient criterion to make a diagnosis of intellectual disability. As it is, the DSM-5 includes the intellectual disability in the section entitled "Intellectual Developmental Disorder," and it categorizes the seriousness of intellectual disability according to four severity levels: mild, moderate, severe, and profound. Moreover, attention is also paid to subjects with borderline intellectual functioning (BIF) and specific learning disorder (SLD) because these conditions correspond to a high level of uncertainty in the diagnostic process (Vianello, Cornoldi, Buono, Termine, & Bartoli, 2017). The level of adaptation is a determining factor to complement the diagnosis, together with the intellectual level (Vianello, Di Nuovo, & Lanfranchi, 2014). As a consequence of the DSM-5 remarks, to assess the intellectual disability and to discriminate it from BIF (a very relevant issue in the diagnostic process for assigning specialized teachers in Italian schools), also the adaptive level has to be assessed. To this end, a range of scales can be considered, Vineland's adaptive behavior scale (VABS; Sparrow, Balla, & Cicchetti, 1984) being the most used. Therefore, diagnoses of intellectual disability are attributed to subjects that reported low scores (below or near the threshold) in both intellectual functions and adaptive functioning.

The WISC-IV Full Scale IQ: A Flattening Evaluation of the Intellectual Functioning

The Full Scale IQ (FSIQ) is considered the most representative estimate of general cognitive ability and it is derived by integrating the results of the following 10 core subtests of the scale (producing factorial indexes): Similarities, vocabulary, and comprehension (verbal comprehension index); block design, picture concepts, and matrix reasoning (perceptual reasoning index); digit span and letter-number sequencing (working memory index); coding and symbol search (processing speed index). At first, partial scores are summed and standardized into IQ points with a mean of 100 and a standard deviation of 15, according to participants' age. Therefore, according to the μ–2σ threshold, the cut-off point for intellectual disability corresponds to the IQ score of 70.

Integrating the four index scores to obtain the FSIQ brings about also the question related to the score variability. In this regard, it is important to consider that the four main WISC-IV indexes are linearly dependent and so the covariance matrix must be estimated to determine the variance of the FSIQ scores. Otherwise, discarding the covariances between index scores, the total variance would be underestimated. To avoid this risk, in the WISC-IV calibration procedure (and the Italian calibration too), the reference values were defined according to the entire covariance matrix. In this way, the variability of the FISQ corresponds to the sum of the variability of all indexes, and the relationships between each pair of indexes are inevitably lost.

Many authors questioned the usefulness of the FSIQ in the diagnostic decision-making process compared to the four index scores (e.g., Fiorello et al., 2007; McGill, 2016). Furthermore, the importance of a complete evaluation of the intellectual functioning based on the four main WISC-IV indexes is also underlined by Hale and Fiorello (2004), who pointed out the inappropriateness of the diagnostic FSIQ score interpretation when there are extreme discrepancies among index scores. In the "Handbook of Psychological Assessment," also Groth-Marnat (2009) warns clinicians against the risk of obtaining impure global measures when differences between the four index scores increase. Similarly, the Technical and Interpretive Manuals of the latest editions of the Wechsler Scales (Wechsler, 2008, 2014) encourage the user to consider the subject's profile based on the main factors, rather than the FSIQ, for their clinical interpretation in the presence of significant differences among index scores. Indeed, considering the WISC-IV scores recorded on subjects that show such significant differences, it results in a distortion of the correlation structure with respect to the normal reference population. Moreover, and more seriously, a complementary effect of the correlation distortion can also be observed when the scores, independently considered, are both below the threshold. Figure 1 shows the latter case. It helps to visualize the risk resulting from the loss of information when the FSIQ alone is considered. It presents the case of a 10-year-old participant who has a score of 19 in both verbal comprehension and perceptual reasoning indexes (VC = 19; PR = 19). The plots consider the two indexes and the respective Gaussian distributions, generated assuming equal marginal distributions. The left-hand plot shows the case of not-correlated index scores, whereas the right-hand plot illustrates positively correlated scores. The correlation value $r = .49$ refers to the parameters found within the Italian calibration of WISC-IV for 10-year-old subjects (Orsini, Pezzuti, & Picone, 2012).

The figure shows that the classification of the same observation changes if we consider the correlation between the index scores. When indexes are independent (Figure 1, left-hand side) the observation falls outside the 95% confidence circle, and the participant should be diagnosed as having an intellectual disability. Instead, in the case of correlated scores (Figure 1, right-hand side), the same observation falls within the 95% confidence ellipse. Therefore, it should be considered as belonging to the normal intellectual functioning group. In a complementary way, it can be argued that participants obtaining a high score in a given intellectual ability index and a very low score in another one, would have been located within the 95% confidence circumference in the first case (independent indexes), while they would fall outside the 95% confidence ellipse in the second case (correlation between index scores). It means that the bivariate approach that includes index score correlation (on the right of Figure 1) could allow us to identify the subjects that fall out of the threshold due to the decline in only one intellectual ability. However, the FSIQ as a single factor embeds the total variance and does not allow the identification of those subjects. Clearly, this is also true when considering more than two scales.

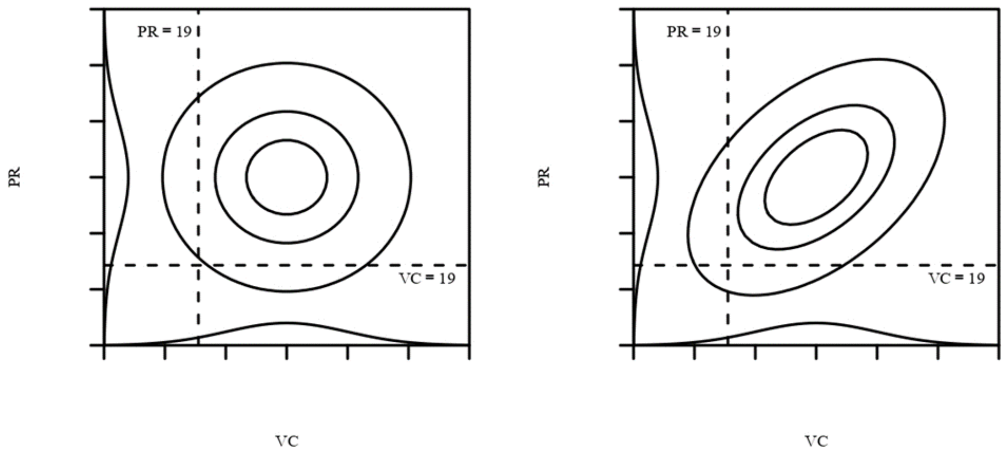

FIGURE 1
Contour plot of a bivariate normal distribution for the vector variable (VC, PR) along with the marginal distribution *f*(VC) and *f*(PR); the dashed lines indicate the case of VC = 19 and PR = 19 while the ellipses correspond to the two-dimensional confidence interval at 95%, 80%, and 68%.
On the left: the VC and PR indexes are considered as independent; on the right: the two index scores have a correlation of $r = .49$.
VC = verbal comprehension; PR = perceptual reasoning.

A multivariate approach, which jointly considers the four dimensions of intellectual functioning, can represent an advance in the evaluation of intellectual functioning. In particular, since we focused on subjects with different intellectual functioning with respect to the four main WISC-IV indexes, we propose Gaussian mixture models (GMMs; McLachlan & Peel, 2000; Titterington, Smith, & Makov, 1985) as a suitable statistical tool to detect differences between subpopulations that have mean and covariance parameters significantly different. Indeed, GMMs allow us to describe populations made up of subpopulations that share the same distribution but have one or more different vectors of parameters and do not assume the local independence of variables. Thanks to their flexibility, GMMs have been extensively used in parametric supervised and not supervised classification problems.

AIM OF THE STUDY

According to the idea that every partial score carries specific information about the intellectual functioning that must be considered for diagnostic purposes, the present study aims to analyze the WISC-IV capability to support the diagnostic decision-making from a multivariate perspective, based on the joint evaluation of the four index scores.

In this work, GMMs describe mixed populations made up of groups of subjects having different intellectual functioning with respect to the four main WISC-IV indexes. The main underlying hypothesis assumes that if the groups have mean and covariance parameters significantly different, GMM should detect these differences.

Therefore, the present work investigates to what extent: (1) the jointly considered WISC-IV indexes can help clinicians to recognize subjects with intellectual disability better than the FSIQ; (2) the normal population multivariate distribution can be defined by a GMM made up by several populations, some of them referring to subjects that present specific intellectual disabilities. Compared to other multivariate methods, GMMs permit to perform multidimensional evaluations taking into account the index score correlations (McLachlan & Peel, 2000).

METHOD

Mixture of Gaussian Distributions

In the field of cluster analysis, mixture models provide a model-based approach to the clustering according to which the *n* observations of the dataset come from *K* different underlying subpopulations that share the same distribution but are defined by different vectors of parameters (McLachlan & Peel, 2000).

Therefore, in its general form, a mixture model is defined by a linear combination of *K* probability density functions, which are called components,

$$f(X|\pi, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k f(X|\boldsymbol{\theta}_k), \tag{1}$$

where X denotes a p-dimensional random variable, K is the number of components; so, each subpopulation can be described through a specific component that has the density function $f(X|\boldsymbol{\theta}_k)$ defined by the parameter vector $\boldsymbol{\theta}_k$in-depth $\forall k \in [1, \dots, K]$. The weight π_k specifies the prior probability of pulling out the k-th component, and it is called mixing coefficient (McLachlan & Peel, 2000). It is worth remembering that the proportions π_k have two constraints (McNicholas, 2016): (*i*) for each component they must be greater than 0 ($\pi_k > 0\ \forall k \in [1, \dots, K]$); (*ii*) their sum must be equal to 1 ($\sum_{k=1}^{K} \pi_k = 1$).

To estimate the mixture component parameters, the component probability distribution family must be known or assumed as known. About that, the most popular choice is to postulate that in each component the random variable is distributed according to the normal distribution (finite GMMs), because Gaussian distribution approximates any sampling distribution according to the statement of the central limit theorem (McLachlan & Basford, 1988; McLachlan & Peel, 2000). If the mixture components are normal, Equation 1 takes the following form:

$$f(X|\pi, \boldsymbol{\mu\Sigma}) = \sum_{k=1}^{K} \pi_k\ N\left(X|\boldsymbol{\mu}_{k,} \boldsymbol{\Sigma}_k\right) \tag{2}$$

where $N\left(X|\boldsymbol{\mu}_{k,} \boldsymbol{\Sigma}_k\right)$ denotes the k-th generic component of the Gaussian mixture distribution and the p-dimensional vector $\boldsymbol{\mu}_k$ and the $p \times p$ covariance matrix $\boldsymbol{\Sigma}_k$, the component parameters (Banfield & Raftery, 1993).

The mixture parameters are usually estimated by maximum likelihood using the expectation-maximization (EM) algorithm, a two-step iteration procedure introduced by Dempster, Laird, and Rubin (1977), (see also McLachlan & Basford, 1988; McNicholas, 2016). Very briefly, in the expectation step, starting from the parameters' initial estimate, the algorithm produces a provisional posterior probability for each observation of belonging to each component. In the maximization step, this provisional posterior probability is used to estimate more accurately the mixture's parameters (Friedman, Hastie, & Tibshirani, 2001).

The procedure iterates until convergence, and each observation $\mathbf{x}_i^T = \left[x_{i1}, x_{i2}, \dots, x_{ip}\right]$ is assigned to the component corresponding to the highest probability of belonging (McLachlan, Peel, & Prado, 1997). For more details about the EM algorithm, see McLachlan and Basford (1988), McLachlan and Peel (2000), Celeux and Govaert (1992).

Furthermore, the large number of parameters to be estimated suggests to impose suitable restrictions on the total number of components and on the covariance matrix structure (e.g., mixture model for probabilistic principal component analyzers; Banfield & Raftery, 1993) to obtain a more parsimonious model that preserves a good adaptation to the data as well (Steinley & Brusco, 2011).

According to Banfield and Raftery (1993), 14 possible models can be estimated based on the following factorization of the covariance matrix $\boldsymbol{\Sigma}_k$:

$$\boldsymbol{\Sigma}_k = \lambda_k \Gamma_k \boldsymbol{\Delta}_k \Gamma_k^{\mathrm{T}} \tag{3}$$

where λ_k is a scalar, Γ_k is a square matrix of the same order of $\boldsymbol{\Sigma}_k$ that defines a basis of vector space, and $\boldsymbol{\Delta}_k$ is a diagonal matrix of the same order of $\boldsymbol{\Sigma}_k$. Each factor affects a geometric characteristic of the *k*-th generic component of the Gaussian mixture component: λ_k the volume, Γ_k the orientation, and $\boldsymbol{\Delta}_k$ the shape. Parsimonious parameterizations can be obtained constraining one or more of these elements and/or assuming that they are equal across groups. For a more detailed description of all the 14 possible models with the corresponding distribution structure type, geometric characteristics and graphical representation in the bi-dimensional case, see Scrucca, Fop, Murphy, and Raftery (2016).

To optimize the model parsimony, and to obtain the optimal model fitting to the data, several statistical criteria exist. The Bayesian information criterion (BIC)[1] is the most widely used index for choosing the best model (McLachlan & Peel, 2000; Schwarz, 1978). The BIC is the Bayesian version of the Akaike information criterion (AIC) that is a parametric measure of the model divergence from the null hypothesis (no subpopulations) penalized by the total number of estimated parameters (Akaike, 1974, 1987). Lower AIC/BIC values correspond to a more informative model.

Mixture models estimate the *a posteriori* probability for each observation to belong to each mixture component. Thus, the probability can be considered a measure of the uncertainty of each observation to belong to a specific component, that is, probability values close to .50 indicate the highest uncertainty. Therefore, each component represents a group of observations whose *within* similarity depends on that uncertainty. In the context of the diagnosis of intellectual disability, this may be the case of subjects with borderline intellectual functioning, which is a condition between normal intellectual functioning and intellectual disability.

It is interesting to note that the underlying probabilistic logic of the mixture models recalls the dimensional approach to psychopathology proposed by the DSM-5; for this reason, multivariate mixture models might help us to overcome the previous dichotomous approach to the psychopathology diagnoses. The dimensional logic relevance in the diagnostic field has also been remarked by Di Nuovo A., Di Nuovo S., Buono, and Cutello (2014), who refer to the specific case of intellectual disabilities. Mixture model's usefulness and potentiality in psychopathology researches are also illustrated in Lenzenweger, Clarkin, Yeomans, Kernberg, and Levy (2008), Lenzenweger, McLachlan, and Rubin (2007), Welham, McLachlan, Davies, and McGrath (2000).

Participants

The study focuses on two groups of participants: 10- and 12-year-old children, respectively. The two studies were conducted both with balanced samples with nearly half of the participants with a clinical diagnosis and the remaining with normal intellectual functioning.

For each age level, a set of 50 participants was randomly generated according to the reference population (at the same age) defined through the Italian calibration of the WISC-IV diagnosis scale (Orsini et al., 2012). It is worth underlining that the choice of generating observations from the distribution of the

Italian WISC-IV reference populations, rather than considering new real participants, allowed us to avoid possible distortions due to a nonrepresentative sampling.

A team of expert psychologists collected the data of participants with intellectual disability at the Research Institute for Mental Retardation and Brain Aging "Oasi Maria SS." in Troina (EN), Sicily, Italy. At the age level of 10 and 12 years, the data respectively contains $N = 52$ and $N = 47$ real-world participants that were classified by the experts in four categories. For age 10 the categories were: $n = 16$ borderline intellectual functioning (BIF), $n = 17$ mild intellectual disability (MiID), $n = 9$ moderate intellectual disability (MoID), and $n = 10$ specific learning disorder (SLD). For age 12, the categories were: $n = 11$ BIF, $n = 16$ MiID, $n = 14$ MoID, and $n = 6$ SLD. These categorical diagnoses were made considering both the intellectual functioning (assessed by the WISC-IV) and the adaptive functioning (assessed by the VABS) with respect to the normative range based on the standardization for the Intellectual Disability of the VABS (Italian version; Balboni & Pedrabissi, 2003), as also reported in Di Nuovo and Buono (2009; Table 1).

Procedure

All analyses were conducted using the R package `mclust` (Scrucca et al., 2016) available at the CRAN repository and running in the R statistical software environment. The WISC-IV four-dimensional multivariate samples, corresponding to the normal population, were sampled from the reference Gaussian distributed population using the R package `MASS` (Ripley et al., 2013). Input parameters were the mean vectors and the variance-covariance matrices of the reference normal Italian populations (10- and 12-year-old). The same sampling procedure was applied to the univariate FSIQ.

The choice of the Gaussian mixture model to describe the whole population was grounded on the WISC scale assumption that the latent variables' score distributions (global scale and main indexes) are well approximated by the Gaussian density function (Orsini et al., 2012; Wechsler, 2003). The input parameters of the model did not include the number of the Gaussian mixture components, according to an exploratory approach. The procedure aims to identify the existence of subpopulations in the data and estimate the correspondent parameter vectors. Such a method chooses the best model according to the BIC (Schwarz, 1978). The `mclust` R package can compute all the possible models according to the constraints imposed on the covariance matrices (Banfield & Raftery, 1993) and the number of the mixture components ranging from 1 to 9, and it selects the model with the lowest BIC.

As the category of each participant is known *a priori*, confusion matrices were computed to compare the classification obtained through the GMM and the classification that corresponds to the clinical diagnosis made by the experts. Confusion matrix counts the number of units that are in the same class corresponding to the *a priori* classification and to the mixture model output.

RESULTS

The findings about 10-year-old participants and 12-year-old participants were comparable. The EM algorithm converged within 50 iterations and the BIC indicated that the best model has two components having different mean vectors, and covariance matrices differing according to a scale parameter. In other words, according to the eigenvalue decomposition of the covariance matrix (Banfield & Raftery, 1993), this means that the two extrapolated components had the same covariance matrix, less than a scale factor. Moreover, the two clusters had the same shape, the same orientation towards the axes, but a different volume. Thus, the four considered variables (i.e., the main WISC-IV indexes) were correlated, and these correlations were the same for the two components. Mixture component details for the two age groups are shown in Table 1.

TABLE 1
Gaussian mixture model: Means, covariance, and correlation matrices for both 10- and 12-year-old group, respectively with *N* = 102 and *N* = 97 participants

			First component mainly formed by participants with normal intellectual functioning					Second component mainly formed by participants with intellectual disability			
			VC	PR	WM	PS		VC	PR	WM	PS
10-year-old	Covariance/ correlation matrix*	VC	68.79	.59	.61	.60	VC	15.71	.59	.61	.60
		PR	34.72	51.65	.42	.49	PR	7.93	11.8	.42	.49
		WM	32.84	19.77	42.11	.56	WM	7.50	4.52	9.62	.56
		PS	30.22	21.28	21.90	36.76	PS	6.90	4.86	5.00	8.40
	Means		27.07	28.76	16.61	17.29		13.46	12.27	5.33	5.52
			VC	PR	WM	PS		VC	PR	WM	PS
12-year-old	Covariance/ correlation matrix*	VC	62.57	.67	.64	.59	VC	17.94	.67	.64	.59
		PR	47.23	79.70	.61	.55	PR	13.54	22.84	.61	.55
		WM	33.80	36.00	44.25	.65	WM	9.69	10.32	12.68	.65
		PS	26.00	27.34	24.37	31.58	PS	7.45	7.84	6.99	9.05
	Means		27.40	28.30	17.62	17.15		10.29	12.20	5.03	5.02

Note. VC = verbal comprehension; PR = perceptual reasoning; WM = working memory; PS = processing speed.
* The diagonal elements are the variances, under them there are the covariances between the variables, and above their correlations.

The procedure identified two subpopulations in the data (see Table 1): The first one mainly consisted of participants with normal intellectual functioning (NF), borderline intellectual functioning (BIF), and specific learning disability (SLD). The second one consisted of participants with intellectual disability, including most of the participants with mild intellectual disability (MiID) and moderate intellectual disability (MoID). The remaining participants showed a posteriori probability for the first component which was very close to $\pi = .50$, denoting a very high uncertainty in classification (this holds for both age groups). Interestingly, most of the participants with MiID, classified in the first component, presented scores very close to the μ–2σ threshold (see Table 2). Moreover, choosing a more restrictive threshold of $\pi = .70$ instead of the standard $\pi = .50$, only seven observations did not exceed it ($n = 2$ for 10-year-old, and $n = 5$ for 12-year-old participants), showing a stable allocation overall. Standard deviations from the normative means of the four WISC-IV index scores of each participant with clinical diagnosis, classified in the first component, are provided in Table 2.

To compare the achieved results with the typical diagnostic decision-making process, in which only the full-scale IQ is considered, we repeated data analyses by using the FSIQ only. For both 10- and 12-year-old participants, the best model selected by BIC had two components, with the following parameters, respectively: $\mu 1 = 102.66$, $\mu 2 = 60.73$, $\sigma = 14.33$; $\mu 1 = 101.90$, $\mu 2 = 56.29$, $\sigma = 14.58$. Notice that the two mixture models have homoscedastic components. The participants' classification based on the FSIQ was different compared to that obtained jointly considering the four scores, especially concerning participants with BIF and SLD. Considering the FSIQ alone, most of the participants with BIF (14 out of 16 for 10-year-old participants and 8 out of 11 for 12-year-old participants) and some of the subjects with SLD (6 out of 10 for 10-year-old participants) were classified in the second component.

TABLE 2
In the table there are the observations with clinical diagnosis classified in the first component (with participants with normal intellectual functioning) by GMM based on the joint consideration of the four main WISC-IV indexes. The clinical diagnosis (Diagn.) and their score difference measured in terms of standard deviations from the reference parameters defined through the Italian calibration of WISC-IV referred to each age group are reported

10-year-old (*N* = 102)						12-year-old (*N* = 97)					
No. observation	Diagn.	VC	PR	WM	PS	No. observation	Diagn.	VC	PR	WM	PS
51	BIF	–1.05	–1.71	–1.02	–1.18	51	BIF	–1.40	–0.87	**–2.67**	–1.37
52	BIF	–1.18	–1.14	**–2.65**	–0.98	52	BIF	–1.53	–0.28	**–2.27**	–1.56
53	BIF	–1.58	–1.29	**–2.86**	–1.18	53	BIF	–0.88	–1.75	–1.04	0.17
54	BIF	–0.66	–1.43	–1.63	–1.78	54	BIF	–1.66	–1.01	–1.24	–0.79
55	BIF	–1.58	–1.29	**–2.65**	–1.78	55	BIF	–0.62	–0.43	**–2.27**	–1.94
56	BIF	–1.71	–1.14	–1.63	–1.38	56	BIF	–0.62	**–2.49**	**–2.67**	**–2.71**
58	BIF	**–2.11**	–1.14	–1.22	–1.58	57	BIF	–0.62	–0.57	**–2.06**	–0.98
59	BIF	–0.53	–0.57	–1.22	**–2.58**	58	BIF	–1.14	–1.31	–1.24	**–2.52**
60	BIF	–1.05	–0.86	0.00	**–2.18**	59	BIF	**–2.18**	–0.43	**–2.06**	–1.37
62	BIF	–0.79	–0.71	–1.63	–1.78	60	BIF	–1.01	–0.43	–1.45	–1.56
63	BIF	–0.53	–0.71	–1.84	–0.58	61	BIF	–0.36	**–2.04**	**–2.27**	–0.79
64	BIF	–1.45	–0.29	–1.02	–1.38	62	MiID	–1.40	**–2.04**	–1.65	–1.17
65	BIF	–1.05	–0.57	**–2.65**	**–2.38**	63	MiID	–1.14	–1.31	–1.86	–1.17
66	BIF	–1.18	–0.57	–0.61	–1.18	65	MiID	–1.27	**–2.04**	**–2.47**	**–2.52**
67	MiID	**–2.24**	–1.29	**–2.86**	**–2.18**	67	MiID	**–2.05**	–1.16	–1.86	–1.75
71	MiID	**–2.24**	–1.43	–1.84	–0.98	72	MiID	–1.40	**–2.04**	–1.65	–1.17
73	MiID	**–2.89**	–1.29	**–3.67**	**–2.58**	73	MiID	–1.66	–1.46	**–2.06**	–1.56
77	MiID	**–2.50**	–1.29	**–3.47**	–1.78	75	MiID	–1.27	**–3.07**	**–2.47**	–1.37
78	MiID	–1.58	–1.29	**–2.86**	**–2.58**	77	MiID	–1.14	–1.01	**–3.08**	**–2.13**
79	MiID	–1.58	–0.71	–1.63	**–2.58**	87	MoID	**–3.22**	**–2.78**	**–3.69**	–1.94
93	SLD	0.00	0.14	–1.43	–1.98						
94	SLD	–0.13	1.14	**–2.45**	0.02						
95	SLD	–0.92	–0.29	**–2.45**	–0.78						
96	SLD	–0.79	–0.29	**–2.65**	**–2.18**						
97	SLD	–0.26	–0.29	–1.63	–1.38						
98	SLD	**–2.11**	0.00	**–2.04**	**–2.58**						
99	SLD	–1.58	0.71	**–2.45**	–1.38						
100	SLD	–0.26	**–2.43**	–1.84	–0.78						
101	SLD	0.00	0.43	–1.02	–0.18						
102	SLD	–0.53	–0.43	–1.43	–0.98						

Note. BIF = borderline intellectual functioning; MilD = mild intellectual disability; SLD = specific learning disability; MoiD = moderate intellectual disability; VC = verbal comprehension; PR = perceptual reasoning; WM = working memory; PS = processing speed. Values in bold are score differences exceded the normative threshold of two standard deviations from the reference parameters.

When considering the correlations between the four WISC-IV index scores, participants with BIF, MiID and, SLD changed their allocation, as hypothesized. As we expected, model classification predictions based on FSIQ were more consistent with clinical psychologist classifications since both used the same criterion variable (i.e., the FSIQ). It is worth remembering that clinicians also considered the adaptive functioning of participants to determine the level of severity of the intellectual disability during the diagnostic process. Furthermore, GMMs with FSIQ better replicated the classification made by clinicians, if we consider only the broad classifications of normal participants against any clinical conditions. When considering the GMM based on the four indexes and their correlations, such a consistency did not occur. These results support our idea that there are marked differences between intellectual functioning evaluation based on the FSIQ and the one grounded on the joint consideration of the four main WISC-IV indexes. Nevertheless, aiming at describing the intellectual functioning in clinical conditions, aside from the broad classification, the four indexes (and their score correlations) permit the identification of more detailed subcategories than the FSIQ can do. Therefore, in our opinion, it is worth examining in depth the characteristics of the GMM component subpopulations, the relationships between the four indexes in each of them, and how this information might contribute to enhance the diagnosis and offer a more detailed description of the intellectual functioning of the subjects.

Some Evidence from SLD and BIF Groups

To study the impact of the observations that mainly affect the correlation structure of the two-component mixture model, the GMMs were estimated removing the SLD and BIF from the dataset.

Since the literature on intellectual and developmental disabilities shows that participants with SLD have different intellectual functioning in relation to the four WISC-IV dimensions, with specific deficiencies in working memory and processing speed (Cornoldi, Giofrè, Orsini, & Pezzuti, 2014; Giofrè & Cornoldi, 2015), we intended to repeat all data analyses dropping the participants with SLD and observing the effects on the new estimated GMM model, and the classification of the subjects. Models related to the 10-year-old participants and the 12-years-old participants were comparable, as observed previously. In both groups, the BIC indicated that the model with two components was the best one and that the two clusters had the same orientation toward the axes, and different volume and shape (the four main WISC-IV indexes had different variances and correlations within the two subpopulations). Table 3 and Table 4 show the two mixture components for the two groups without SLD and the confusion matrices, respectively.

It is worth noting that these results, obtained discarding SLD participants from the analysis, lead us to argue that participants with SLD were classified within the first component, in the whole sample analysis, mainly because their mean scores are similar to the normal intellectual functioning individuals.[2] Indeed, in the second analysis (without SLD participants), the results about the correlation matrix suggest that SLD participants' correlation matrix is similar to the one of participants with intellectual disabilities. Thus, we can conclude that in the whole sample model, within component correlation matrices were assumed to be equal, because the participants with SLD were included.

Therefore, in this specific case, considering mixture components having two different correlation matrices is more appropriate than considering a common matrix. Indeed, when the model had two components with different correlation matrices, it retraced the group of participants with normal intellectual functioning with greater precision.

By also removing the BIF from the dataset, as expected, the results were comparable to the ones obtained removing only participants with SLD, both regarding the best model selected and the classification of observations. The differences between the correlation matrices remained almost the same, and the correlations between the index scores in the group of participants with intellectual disability were higher than those in the group of participants with normal intellectual functioning. This further result argued in favor of mixture components having different correlation matrices rather than a single common correlation matrix.

TABLE 3
GMM after removing participants with SLD: Means, covariance, and correlation matrices for both 10- and 12-year-old group

			First component mainly formed by participants with normal intellectual functioning				Second component mainly formed by participants with intellectual disability				
			VC	PR	WM	PS		VC	PR	WM	PS
10-year-old	Covariance/ correlation matrix*	VC	55.67	.56	.43	.44	VC	27.39	.49	.77	.62
		PR	27.30	43.18	.36	.41	PR	12.26	23.27	.30	.68
		WM	17.04	12.69	28.62	.33	WM	12.60	4.53	9.69	.53
		PS	19.05	15.47	10.33	33.58	PS	10.78	10.96	5.46	11.09
	Means		28.86	29.79	19.00	18.82		14.18	14.58	5.52	6.75
			VC	PR	WM	PS		VC	PR	WM	PS
12-year-old	Covariance/ correlation matrix*	VC	72.30	.66	.45	.40	VC	22.62	.60	.76	.54
		PR	43.25	58.69	.47	.42	PR	12.84	20.54	.56	.45
		WM	24.33	22.90	39.92	.43	WM	10.93	7.59	9.03	.30
		PS	17.86	16.91	14.24	27.45	PS	7.23	5.78	2.55	7.87
	Means		27.57	27.58	17.20	17.11		10.46	12.17	4.90	5.00

Note. SLD = specific learning disability; VC = verbal comprehension; PR = perceptual reasoning; WM = working memory; PS = processing speed. * The diagonal elements are the variances, under them there are the covariances between the variables, and above their correlations.

TABLE 4
Confusion matrices show a comparison between clinical psychologists' diagnoses and GMMs' classification for both age groups with and without SLD participants

	10-year-old					12-year-old			
	N = 102		Without SLD (*N* = 92)			*N* = 97		Without SLD (*N* = 91)	
	Comp. 1	Comp. 2	Comp. 1	Comp. 2		Comp. 1	Comp. 2	Comp. 1	Comp. 2
NF	50	0	50	0	NF	50	0	50	0
BIF	14	2	10	6	BIF	11	0	11	0
SLD	10	0	-	-	SLD	6	0	-	-
MiID	6	11	2	15	MiID	8	8	7	9
MoID	0	9	0	9	MoID	1	13	0	14

Note. NF = normal intellectual functioning; BIF = borderline intellectual functioning; SLD = specific learning disability; MilD = mild intellectual disability; MoiD = moderate intellectual disability; Comp = component of the mixture.

Further Evidence from Supervised Classification Methods

Unsupervised GMM showed that the WISC-IV led to different results when the four indexes were separately taken into account, rather than the FSIQ. So, to get more evidence, we carried out a supervised classification analysis in which the initial classification (subjects with normal intellectual functioning vs. subjects with clinical diagnosis) was provided as a dependent variable. The main goal was to explore the GMM performance in a supervised context and compare the results with the unsupervised conditions. In particular, we aimed at assessing the predictive capabilities of the four WISC-IV indexes and the FSIQ with respect to the classification provided by the clinicians. For these purposes, the following supervised classification methods were considered: logistic regression (considering both the FSIQ and the four main WISC-IV indexes jointly) and discriminant analysis based on Gaussian finite mixture modeling (Friedman et al., 2001). The discriminant analysis assumed that, in each group, the distribution function of the variables is a single component multivariate Gaussian (Friedman et al., 2001). A cross-validation procedure was carried out for the two methodologies, in both age groups, to obtain an unbiased estimation of the classification error rate.

It is worth remembering that in the multivariate case, the logistic regression assumes that the variables are locally independent while the discriminant analysis also considers their correlations.

In both age groups, the BIC indicated that the group of subjects with normal intellectual functioning was best described by a spherical multivariate normal while the group of subjects with clinical diagnosis by an ellipsoidal multivariate normal. As we expected and as can also be seen from the confusion matrices in Table 5, the supervised methods provided classifications more similar to the clinician ones compared to the unsupervised case. Furthermore, the supervised methodologies showed approximately the same result in terms of correct classification, with a slightly worse performance in the discriminant analysis in which the four indexes and their correlations were considered. It is worth noting that in both cases, the incorrectly classified clinical observations were SLD and BIF participants, as a confirmation of the uncertainty about these clinical categories.

It is also interesting to note that the observations incorrectly classified by the discriminant analysis included almost all the incorrectly classified observations by the four-dimension logistic regression and the logistic regression with the FSIQ.[3]

Moreover, we detected that the few additional observations that the discriminant analysis wrongly classified with respect to the four-dimension logistic regression are participants with SLD that fall outside the confidence ellipse of the score distribution of participants with normal intellectual functioning if the main WISC-IV indexes are independently considered. On the contrary, they fall inside when the index scores are considered correlated. So, it is possible to figure that these participants represent conditions such as the ones described in Figure 1. It is worth noting that if we consider subjects with SLD having a normal general intellectual functioning (American Psychiatric Association, 2013), the classification performance of the discriminant analysis based on Gaussian finite mixture modeling became better than the other methods.

This further in-depth analysis underlines, on the one hand, the adequacy of GMMs in describing the latent structure in the data: It gave results comparable to those of four-dimensional logistic regression except for few subjects. On the other hand, the GMM might lead to include or exclude observations from a given class according to the scores obtained in each single index and on the evaluation of index score correlations. This issue mainly affects BIF and SLD subjects who have different intellectual functioning concerning the four WISC-IV dimensions, as described in the literature (Cornoldi et al., 2014; Giofrè & Cornoldi, 2015). As a confirmation of the evidence obtained by mixture models in unsupervised conditions, the differences between logistic regression and discriminant analysis results indicate the presence of peculiar relationships between the main four WISC-IV indexes in participants with SLD that should be explored in depth.

TABLE 5
Confusion matrices about the logistic regression (considering both the FSIQ and the four main WISC-IV indexes) and the discriminant analysis based on Gaussian finite mixture modeling.
On the horizontal lines there are the real classifications (denoted by the subscript "r"), on the vertical lines the models predictions (denoted by the subscript "p").

Logistic regression (Full Scale IQ)						Logistic regression (multiple regression)						Discriminant analysis (correlated indexes)					
10-year-old			12-year-old			10-year-old			12-year-old			10-year-old			12-year-old		
	N_p	D_p		N_p	D_p		N_p	D_p		N_p	D_p		N_p	D_p		N_p	D_p
N_r	47	3	N_r	47	3	N_r	47	3	N_r	46	4	N_r	47	3	N_r	45	5
D_r	3	49	D_r	4	43	D_r	2	50	D_r	3	44	D_r	5	47	D_r	6	41

Note. FSIQ = full-scale intelligent quotient; WISC-IV = Wechsler Intelligence Scale for Children-IV; N = subjects with normal intellectual functioning; D = subjects with clinical diagnosis.

DISCUSSION OF THE RESULTS

This study aimed to examine the WISC-IV capability to support the diagnostic decision-making process in a multidimensional approach based on the joint evaluation of the four main index scores. We started from the idea that every partial score carries a specific information about the intellectual functioning, that must be taken into consideration, and we were worried about the risk of using an inappropriate FSIQ global measure, especially when differences between index scores increase (Groth-Marnat, 2009; Hale & Fiorello, 2004; Wechsler, 2008; 2014). Therefore, we proposed the GMM as a statistical tool to cope with those issues (McLachlan & Basford, 1988; McLachlan & Peel, 2000; McNicholas, 2016).

Two main advantages derive from the use of the mixture models: (i) the multivariate mixture models allow adopting a multidimensional approach; and (ii) mixture models do not assume local independence between the variables so that a specific covariance matrix can characterize each cluster. Indeed, considering the index score correlations, the confidence boundaries for the diagnosis definition were modified.

Results suggested that individuals with intellectual disabilities presented a specific distribution of IQ values and specific correlations between the variables that describe the intellectual functioning. In particular, results emphasized that considering individuals with intellectual disability as merely belonging to the tails of the normal intellectual functioning distribution causes an unfair flattening in the definition of their intellectual functioning. Indeed, they present their own specific characteristics, as also suggested in the DSM-5 (American Psychiatric Association, 2013).

Our results demonstrated that models that do not consider the correlation structures among the WISC-IV score indexes can accurately separate normal and clinical participants (see Table 5).

Furthermore, about the WISC-IV capability in differentiating among individuals with different intellectual functioning, it can be seen that the joint consideration of the four main WISC-IV indexes enhances the overall specificity. Indeed, results indicate that SLD participants are classified as normal since they have just one score index greater than 2 standard deviations, as seen in Table 2. This is consistent with the DSM-5 policy that considers SLD as subjects with normal intellectual functioning, even if they are affected by learning disorders. Moreover, removing the SLD category form the analysis the sensitivity increases for the 10-year-old group (the number of wrongly classified MiID subjects decreases from six to two). As a consequence, a more accurate diagnosis of subjects with BIF and SLD needs to consider additional dimensions, as suggested by the DSM-5 (American Psychiatric Association, 2013).

Therefore, a multidimensional approach such as the mixture models might help us to have a more detailed preliminary classification, especially for diagnostic conditions characterized by a higher uncertainty level, such as subjects with SLD and BIF.

Additionally, since mixture models give us the degree of classification uncertainty of each observation, they help to identify the observations for which any group may appear to be not entirely appropriate. For example, this was the case of some observations that were wrongly classified on the basis of a probability value π close to .50, in which the allocation of the observation can be considered highly uncertain.

Further evaluations concerning the relationship between the indexes involved in the assessment process were made thanks to the use of the mixture models in an unsupervised condition that shed light on some specific characteristics of the categories.

Conclusion and Future Perspectives

The clinical practice, the DSM-5 indications, and the discussed results showed that the diagnosis of intellectual disability could hardly be accomplished only by assuming the non-compliance with the so-called “norm.” In particular, there is an area in which any diagnosis could be inappropriate. Specifically, it is speaking about subjects with BIF, which is a condition between normal intellectual functioning and intellectual disability. Subjects with this type of intellectual functioning are located close to the cut-off point that, in cases like these, assumes an even more decisive role. Furthermore, also subjects with SLD showed considerable uncertainty in their allocation. In this sense, the posterior probability of belonging to the components of the mixture, which is provided by the EM algorithm during the estimation of the mixture model, would help us. In further, after collecting more data to achieve higher stability of the obtained results, we could establish a rule according to which the classification of an observation can be considered secure if the corresponding probability of belonging to the component of the mixture overcomes a fixed threshold (for example $\pi = .70$), otherwise the observations will be subjected to further evaluations.

Another interesting solution consists in considering additional variables that influence the intellectual functioning in order to integrate the results. The goal of this procedure is to reduce the evaluation errors by building variable thresholds on the WISC-IV global scale that will vary according to the different values obtained by the participants in other tests, for instance, in the adaptive functioning scales, as the DSM-5 suggests.

Another interesting aspect arising from our study, that should be further explored, is related to participants with SLD. The study demonstrates the peculiarity of their intellectual profile that must be differentiated both from intellectual disability and from normal intellectual functioning. Other authors also used the WISC-IV scores to show evidence regarding this need, for example, Cornoldi et al. (2014), Giofrè and Cornoldi (2015). Our suggestion is in line with their concern about the inadequacy of the FSIQ as a good measure of intellectual functioning for subjects with SLD.

Finally, it is important to consider that our results were obtained taking into account a small number of observations and thus, further validations would be needed. Starting from the idea that GMMs can improve diagnostic assessments, we are encouraged to carry out more studies in this direction, aimed at defining more accurately the parameters of the WISC-IV subpopulations. In particular, the scores of subjects with intellectual disabilities could be described in a more reliable way to overcome the flattening of these subjects’ profiles.

NOTES

1. As the BIC is an information criterion, the model presenting the lowest BIC is the most informative one. However, a heuristic approach can help the researcher in this choice by taking into account that: a difference in the BIC value less than 2 is considered an irrelevant improvement; a difference between 2 and 6 indicates a modest improvement; a difference between 6 and 10 represents a significant improvement; and, finally, a difference greater than 10 describes a remarkable strong improvement (Cavanaugh, 2016).
2. Independent samples *t*-tests were used to examine the differences in the four main WISC-IV index scores between participants with SLD in our samples and the normative sample recognized as normal intellectual functioning in the Italian calibration of WISC-IV. Analyses highlighted a statistically significant difference only for working memory (WM) and processing speed (PS) indexes: $t_{(208)} = -10.24$ ($p < .001$) and $t_{(208)} = -4.40$ ($p < .001$) for 10-year-old participants, and $t_{(204)} = 4.81$ ($p < .001$) and $t_{(204)} = -3.30$ ($p < .01$) for 12-year-old participants. As shown in the literature, participants with SLD present specific deficiencies in WM and PS (Giofrè & Cornoldi, 2015). However, it is worth noting that these univariate analyses reduce the considered information because the covariance matrix was discarded.
3. The incorrectly classified observations for the 10-year-old group are: nos. 19-40-42 (NF) and 93-94-101 (SLD), for the logistic regression with the FSIQ; nos. 25-33-43 (NF) and 66 (BIF) and 101 (SLD), for the four-dimension logistic regression; nos. 19-25-33 (NF) and 60 (BIF) and 93-94-100-101 (SLD), for the discriminant analysis. On the other hand, the incorrectly classified observations for 12-year-old group are: nos. 19-40-42 (NF) and 92-93-94-95 (SLD), for the logistic regression with the FSIQ; nos. 11-14-19-45 (NF) and 53 (BIF) and 93-97 (SLD), for the four-dimension logistic regression; nos. 11-14-19-41-45 (NF) and 53 (BIF) and 92-93-94-95-97 (SLD), for the discriminant analysis.

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). http://doi.org/10.1176/appi.books.9780890425596

American Psychiatric Association. (2000). *DSM-IV-TR: Diagnostic and statistical manual of mental disorders* (4th ed.; Text Revision). Washington, DC: American Psychiatric Association.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723. http://doi.org/10.1109/TAC.1974.1100705

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*(3), 317-332. http://doi.org/10.1007/BF02294359

Balboni, G., & Pedrabissi, L. (2003). *Adattamento italiano delle Vineland Adaptive Behavior Scales* [Italian adaptation of Vineland Adaptive Behavior Scales]. Firenze, Italy: Giunti OS.

Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, *49*(3), 803-821. http://doi.org/10.2307/2532201

Cavanaugh, J. E. (2016). Model Selection: Bayesian Information Criterion. *Wiley StatsRef: Statistics Reference Online*. http://doi.org/10.1002/9781118445112.stat00247.pub2

Celeux, G., & Govaert, G. (1992). A classification EM algorithm for clustering and two stochastic versions. *Computational Statistics & Data Analysis*, *14*(3), 315-332. http://doi.org/10.1016/0167-9473(92)90042-E

Cornoldi, C., Giofrè, D., Orsini, A., & Pezzuti, L. (2014). Differences in the intellectual profile of children with intellectual vs. learning disability. *Research in Developmental Disabilities*, *35*(9), 2224-2230. http://doi.org/10.1016/j.ridd.2014.05.013

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1-38.

Di Nuovo, A., Di Nuovo, S., Buono, S., & Cutello, V. (2014). Benefits of fuzzy logic in the assessment of intellectual disability. In *Proceedings of the 2014 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE* (pp. 1843-1850). http://doi.org/10.1109/FUZZ-IEEE.2014.6891834

Di Nuovo, S., & Buono, S. (2009). Diagnosis of intellectual disability: Comparison between clinical criteria and automatized procedures. *Life Span and Disability*, *12*(2), 151-159.

Evers, A., Muñiz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J. R., . . ., Urbànek, T. (2012). Testing practices in the 21st century. *European Psychologist*, *17*(4), 300-319. http://doi.org/10.1027/1016-9040/a000102

Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant?. *Applied Neuropsychology*, *14*(1), 2-12. http://doi.org/10.1080/09084280701280338

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). New York, NY: Springer. http://doi.org/10.1007/978-0-387-84858-7

Giofrè, D., & Cornoldi, C. (2015). The structure of intelligence in children with specific learning disabilities is different as compared to typically development children. *Intelligence*, *52*, 36-43. http://doi.org/10.1016/j.intell.2015.07.002

Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). Hoboken, NJ: John Wiley.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. New York, NY: Guilford.

Lang, M., Di Pierro, P., Michelotti, C., & Squarza, C. (2017). *WISC-IV. Wechsler Intelligence Scale for Children: Lettura dei risultati e interpretazione clinica* [WISC-IV. Wechsler Intelligence Scale for Children: Analysis of results and clinical interpretation]. Milano, Italy: Raffaello Cortina Editore.

Lenzenweger, M. F., Clarkin, J. F., Yeomans, F. E., Kernberg, O. F., & Levy, K. N. (2008). Refining the borderline personality disorder phenotype through finite mixture modeling: Implications for classification. *Journal of Personality Disorders*, *22*(4), 313-331. http://doi.org/10.1521/pedi.2008.22.4.313

Lenzenweger, M. F., McLachlan, G., & Rubin, D. B. (2007). Resolving the latent structure of schizophrenia endophenotypes using expectation-maximization-based finite mixture modeling. *Journal of Abnormal Psychology*, *116*(1), 16-29. http://doi.org/10.1037/0021-843X.116.1.16

McGill, R.J. (2016). Invalidating the full scale IQ score in the presence of significant factor score variability: Clinical acumen or clinical illusion?. *Archives of Assessment Psychology*, *6*(1), 49-79.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering* (Vol. 84). New York, NY: Marcel Dekker. http://doi.org/10.1191/0962280204sm372ra

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley-Interscience.

McLachlan, G. J., Peel, D., & Prado, P. (1997). Clustering via normal mixture models. *Proceedings of the American Statistical Society*, 98-103.

McNicholas, P. D. (2016). *Mixture model-based classification*. New York, NY: CRC Press. http://doi.org/10.1007/s00357-016-9211-9

Orsini, A., Pezzuti, L., & Picone, L. (2012). *WISC-IV: Contributo alla taratura Italiana* [WISC-IV: Contribution to the Italian calibration]. Firenze, Italy: Giunti O. S.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'MASS'. Retrieved from CRAN: http://cran.r-project.org/web/packages/MASS/MASS.pdf

Steinley, D., & Brusco, M. J. (2011). Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods*, *16*(1), 63-79. http://doi.org/10.1037/a0022673

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461-464. http://doi.org/10.1214/aos/1176344136

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R journal*, *8*(1), 289-317. http://doi.org/10.32614/RJ-2016-021

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland adaptive behavior scales*. Circle Pines, MN: American Guidelines Services.

Titterington, D., Smith, A., & Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York, NY: Wiley.

Vianello, R., Cornoldi, C., Buono, S., Termine, C., & Bartoli, B. (2017). Riflessioni sul funzionamento intellettivo limite o borderline: proposta per una discussione. Interventi di Serafino Buono, Cristiano Termine e Beatrice Bartoli [Reflections on the borderline intellective functioning: Proposal for a discussion. Speeches by Serafino Buono, Cristiano Termine and Beatrice Bartoli]. *Psicologia clinica dello sviluppo*, *21*(2), 257-276. http://doi.org/10.1449/87537

Vianello, R., Di Nuovo, S., & Lanfranchi, S. (2014). *Bisogni Educativi Speciali: il Funzionamento Intellettivo Limite o Borderline* [Special Educational Needs: Borderline Intellectual Functioning]. Azzano San Paolo, Italy: Junior.

Wechsler, D. (2003). *Wechsler intelligence scale for children-WISC-IV*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale - Fourth Edition technical and interpretive manual*. San Antonio, TX: NCS Pearson.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children - Fifth Edition technical and interpretive manual*. San Antonio, TX: NCS Pearson.

Welham, J., McLachlan, G., Davies, G., & McGrath, J. (2000). Heterogeneity in schizophrenia; mixture modelling of age-at-first-admission, gender and diagnosis. *Acta Psychiatrica Scandinavica*, *101*(4), 312-317. http://doi.org/10.1034/j.1600-0447.2000.101004312.x