Life Span and Disability XII, 2 (2009), 151-159

Diagnosis of Intellectual Disability: comparison between clinical criteria and automatized procedures

Santo Di Nuovo[1] *& Serafino Buono*[2]

Abstract

In a previous work, published in this Journal (n.10/2), we outlined the utility to find shared criteria for diagnosing the levels of Intellectual Disability / Mental Retardation. We presented a computerized algorithm based on the predefined and standardized integration of scores derived from intellectual and adaptive tests.
In this paper we will report the results of the comparative analysis between the commonly used clinical diagnoses and those derived from the automatized procedure, applied to the same cases.
The study was conducted on a sample of 100 diagnoses, regarding different levels of Mental Retardation (n=56), Borderline Intellective Functioning (BIF, n=25), and cases without disability (n=19).
The comparison of the two modality of diagnosis was performed through factorial analysis of correspondences. The analysis showed that the clinical diagnosis an that based on the computerized algorithm agree with respect to the principal dimensions explaining the evaluation criteria.
The overall concordance is 68%, with higher peak for the normal subjects and lower for the BIF, resulting the more complex diagnosis.
The main differences were object of an in-depth analysis, at a qualitative level, examining the profiles of discrepant cases, regarding always adjacent levels.
It was further verified how much the diagnosis was related to the only value of IQ, obtaining a correlation (Spearman's r_S) of .44 with the clinical diagnosis and .47 with the computerized one.
In conclusion, the standardized algorithm applied by the software supplies a

* *Received*: 31 July 2009, Revised: 20 October 2009, Accepted: 20 October 2009.

[1] University of Catania and Enna 'Kore', Italy, e-mail: s.dinuovo@unict.it
[2] IRCCS 'Oasi', Troina, Italy, e-mail: fbuono@oasi.en.it

diagnosis largely overlapping with that based on the clinical evaluation, but with differences justifying the comparative use of the different approaches.

Keywords: Intellectual Disability, Mental Retardation, Borderline Intellectual Functioning, Psychometric evaluation

1. Introduction

The evaluation of the existence and the degree of Intellectual Disability has to be based on internationally shared criteria, according to the DSM-IV-TR (A.P.A., 2000) and ICD-10 (W.H.O., 1992) definitions and to the more recent acquisitions (Luckasson, Borthwick-Duffy, Buntx, Coulter, Craig, Reeve *et al.,* 2002).

To apply these criteria, the integration of psychometric and clinical methods needs a complex mediation. So, it may be useful to make homogeneous this evaluation using standardized algorithms.

The clinical diagnosis of Intellectual Disability (Schalock & Luckasson, 2005) takes into account, besides the Intellectual Quotient, also the evaluation of other relevant variables, deducible from information about subject's behavior and from reports by parents and/or personnel about the overall functioning of the examined person.

In the diagnosis of Intellectual Disability made at the Institute IRCCS at Troina – as in other centres specialized for this kind of diagnosis – the evaluation synthesizing the available elements is made by a diagnostic équipe composed of specialized physicians, psychologists, pedagogists, social workers, interacting with families.

This approach takes into account, besides IQ data, also variables not quantifiable through psychometric instruments, qualitative observations referring to social context, to its attitude toward disability, to the available support; moreover, to the developmental aspects of the deficit and its consequences, and – last but not least for importance – to the affective, emotional and motivational features fully understandable at clinical interview.

The clinical evaluation has to be integrated with the psychometric one, according to already quoted APA and WHO criteria, with a final mediation if these evaluations are discordant, e.g. the tests underscore the person's real capacity of social relation (referred by parents or directly observed), or at the contrary, the intellectual test scores are surprisingly higher than expected on the basis of the social adjustment the subject, due to his/her comorbid disturbances.

We implemented in a software an algorithm, based on artificial neural networks, widely described in Aa. Vv. (2007), which makes in a standardized way the procedure of integration between the data of intellectual tests (IQ Wechsler is applicable, or an equivalent score deducted from other

tests[3]) and the scaled scores obtained in the *Vineland Adaptive Behavior Scale* (*VABS*), with reference to *communication* and *socialization* areas.

The algorithm is based therefore exclusively on psychometric tests scores (one of these based on observation), integrating them in a rigorously standardized way and avoiding 'subjective' evaluations on the part of the personnel.

The schema used for the diagnosis through the computerized algorithm may be synthesized as follows (table 1).

Table 1 - *Schema for the diagnosis integrating IQ and index of adjustment*

Based on total IQ *(Wechsler or derived by other tests):*	***Based on adjustment*** *Position on the VABS normative range*[1]		
	Higher	*Equal*	*Lower*
86-90 Normal	Normal	Normal	Normal
76-85 Borderline Intellectual Functioning	Normal	BIF	BIF
70-75 Borderline Intellectual Functioning[2]	BIF	BIF	Mild
56-69 Mild Mental Retardation	BIF	Mild	Mild
50-55 Mild Mental Retardation	Mild	Mild	Moderate
41-49 Moderate Mental Retardation	Mild	Moderate	Moderate
35-40 Moderate Mental Retardation	Moderate	Moderate	Severe
26-34 Severe Mental Retardation	Moderate	Severe	Severe
20-25 Severe Mental Retardation	Severe	Severe	Profound

[1] The criteria to locate the diagnosis with respect to the normative range are based on the standardization for the Intellectual Disability of the *Vineland Adaptive Behavior Scale*. As cut-off was used the subdivision in three levels presented in the VABS Manual (Sparrow, Balla, & Cicchetti,1984, Italian edition 2003).

[2] According to DSM-IV-TR (A.P.A., 2000) diagnostic criteria, it is possible to make diagnosis of Mental Retardation in persons "with IQ between 70 and 75, who exhibit significant deficits in adaptive behavior. Conversely, Mental Retardation would not be diagnosed in an individual with an IQ lower than 70 if there are no significant deficits or impairments in adaptive functioning." (p. 42).

The aim of the research presented here is to validate this integrated modality of diagnosis, comparing it in a blinded way with the diagnosis independently made by the diagnostic équipe, according to the above quoted criteria, and with that obtained on the basis of only IQ.

[3] The tests chosen to change scores in equivalent IQ, when Wechsler Scale is not applicable, are: *Coloured Progressive Matrices* test (CPM), *Leiter International Performance Scale* (LIPS), Griffiths' *Mental Development Scale* (MDS), *Psycho Educational Profile Revised* (PEP) for autism. The tests, and the reasons of the choice, are described in detail in the previously quoted article (Aa. Vv., 2007).

2. Sample

The cases included in the study were 100, randomly chosen among those diagnosed in one-year period in the IRCCS 'Oasi' Institute.

The cases with Intellectual Disability was divided in the following groups on the basis of the diagnosis included in the clinical record:

- Mild Mental Retardation: n=34
- Moderate Mental Retardation: n= 21
- Severe Mental Retardation n=1 (in this only case a test of intelligence was applicable; other cases of this level, or cases of Profound Mental Retardation, have not been included, due to the unavailability of cognitive tests useful for the computerized diagnosis)
- Borderline Intellectual Functioning: n=25

N=19 cases without any disability were added. They had been submitted to the équipe for learning disabilities but they did not fit the criteria for the diagnosis of Borderline Intellectual Functioning.

The diagnoses had been made independently by the équipe according with the usual criteria, and have been reformulated by means of the software applying the specific algorithm.

3. Analysis of data

The raw correlation between the two evaluations, computed by means of the Spearmans' rank coefficient r_S, after the co-graduation of the levels, is .64. This first overall comparison shows a wide but not complete overlapping of the different classification criteria.

A Correspondence analysis, based on a two-way table (Greenacre, 1984), was performed. The diagnostic groups with a sufficient number of cases were considered, excluding the 'Severe Mental Retardation' diagnosis, due to very small number of cases: only one according to the clinical evaluation, two according to the computerized algorithm.

The aim of Correspondence analysis is to examine the relationship between categorical variables, decomposing contingency tables into row and column coordinates, which are displayed in a graph. Categories that are similar to each other appear close to each other in the graph, representing the principal dimensions. The frequencies related to each categories (*mass*), the indices of variability (*inertia*), and an equivalent of *communality* in factorial analysis (here called *quality*) are computed.

Table 2 synthesizes the results of the analysis, while the graphic representation is shown in figure 1.

The values related to the two diagnostic modalities are very similar and in some cases identical: i.e., in the graph the points for the coordinates pertaining to the different diagnostic groups are quite completely overlapping.

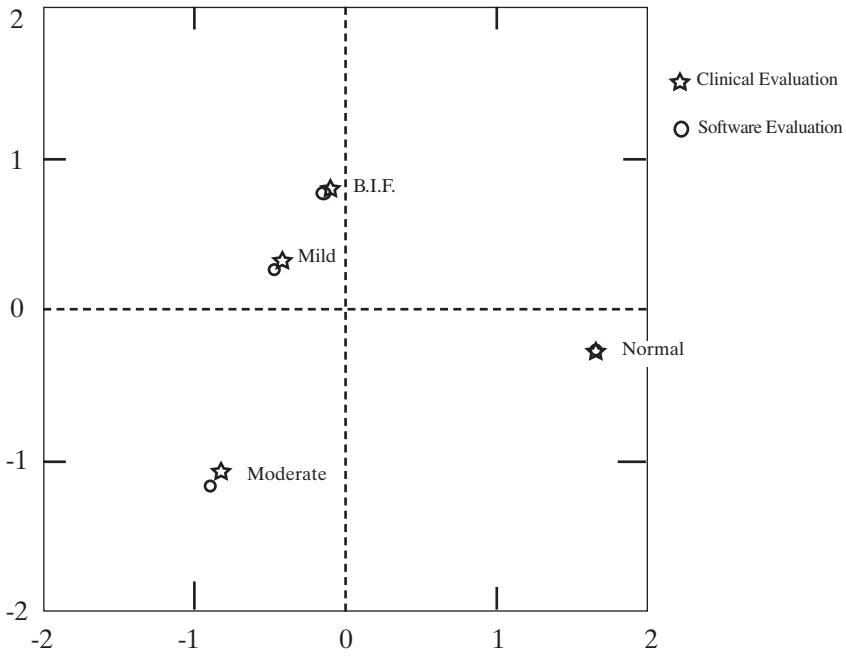
We may conclude that substantially the clinical diagnosis and that based on the computerized algorithm agree with respect to the principal dimensions explaining the evaluation criteria.

Table 2 - *Results of correspondence analysis. Differences in diagnostic categories between clinical and software evaluation*

	Mass		**Quality**		**Inertia**		**Factor 1**		**Factor 2**	
	Clin	*Softw*	*Clin*	*Softw*	*Clin*	*Softw*	*Clin*	*Softw*	*Clin*	*Softw*
Normals	0.19	0.19	1.00	1.00	0.61	0.62	1.73	1.74	-0.40	-0.40
BIF	0.26	0.27	0.86	0.86	0.18	0.18	-0.04	-0.05	0.78	0.76
Mild	0.35	0.32	0.77	0.76	0.14	0.13	-0.45	-0.44	0.31	0.36
Moderate	0.20	0.22	0.99	0.99	0.40	0.41	-0.83	-0.82	-1.12	-1.06

General Chi-square = 130,90, g.l. 9, p<.001; total inertia = 1.34

Figure 1 - *Plot of diagnostic categories in the two factorial dimensions emerging from correspondence analysis*



To evaluate in a more analytical way the discrepancies, we can directly inspect the contingency table crossing the two diagnoses for the same subjects. In this case the diagnosis of 'Severe Mental Retardation' is considered too. The table 3 shows concordances and differences, along with a summary of the percentages of congruent diagnostic categorization.

These percentages are generally high, although few cases of discrepancies are registered, regarding adjacent levels, never more than one level of difference. The overall percentage of concordance is 68 out of 100 cases, with higher peak for normal subjects (89.5%) and lower for BIF (60.0%) and Mild MR (55.9%), diagnostic categories being – as all clinicians know well – the most complex and difficult to discriminate.

Table 3 - *Contingency table between clinical evaluations (rows) and software (columns). Percent of congruent classifications by rows and columns*

	Diagnosis through software						
Clinical Diagnosis	*Normals*	*BIF*	*Mild MR*	*Moderate MR*	*Severe MR*	***Total***	*%*
Normals	**17**	2	0	0	0	**19**	*89.47*
BIF	2	**15**	8	0	0	**25**	*60.00*
Mild MR	0	9	**19**	6	0	**34**	*55.88*
Moderate MR	0	0	4	**16**	1	**21**	*76.19*
Severe MR	0	0	0	0	**1**	**1**	*100.00*
Total	**19**	**26**	**31**	**22**	**2**	**100**	
%	*89.47*	*57.69*	*61.29*	*72.73*	*50.00*		*68.00*

We found it interesting to further elaborate, at a qualitative level of analysis, where the main diagnostic differences are located, and what the profile of the discrepant cases are (evidenced in the table).

Table 3 shows that the most frequent divergences are found between the diagnoses of BIF and Mild MR (17% out of the total). In 9 cases the diagnosis indicated a superior level using the algorithm, in 8 cases the contrary happens.

Six cases of divergence regard the diagnoses between Mild MR and Moderate MR, clinical diagnosis being the more favorable, while in other 4 cases the opposite is found.

We focused on these discrepancies (others being very marginal) the qualitative analysis of the profiles in the subgroups where the diagnostic discordance was verified.

As regards the chronological age, the cases in which the algorithm overvalues the level of the clinical diagnosis – attributing the BIF level instead of Mild MR - result to have a mean age slightly inferior with respect to the cases for which the contrary happens (114.67 months vs 164.88).

The cases, too, where the diagnosis of Mild vs Moderate MR is more favourable if it is made using the software, have a lower age: 168.25 months vs 249.17 of the cases where the clinical diagnosis is more favorable. For both the comparisons the difference is not statistically significant (Mann-Whitney U test, $p>.05$): although a slight bias linked to age may be hypothesized, the discrepancies between the evaluations cannot be considered different from random ones.

It was interesting to repeat the same comparison with reference to Wechsler IQ.

The mean of the group to whom the algorithm assigned a BIF level, while a Mild MR was diagnosed by the clinical evaluation, is 62.78 against a mean of 57.12 for the cases assigned by clinical assessment to BIF instead Mild MR. This difference is statistically significant (Mann-Whitney U test, $p<.05$).

The same trend, although not reaching the significance threshold, is registered regarding the discrepancies between the diagnoses of Mild and Moderate MR; i.e., the cases with a more favourable computerized diagnosis have an higher mean IQ (43.25) than the cases for whom the opposite is true (39.83).

The differences between the groups with disagreeing diagnoses are very limited with respect to the two variables of the Vineland Adaptive Scale entering in the computerized algorithm. The social adjustment, as measured by the test, seems not to influence the diagnostic discrepancies.

To answer the second question posed in our study, we tested how much the diagnosis is related to the IQ alone.

The classification based only on IQ value (according to the criteria of correspondence between IQ and levels shown in the left column of table 1) shows a Spearman's r_S rank coefficient .44 with clinical diagnosis and .47 with the computerized one. The overlapping is wide, therefore is more relevant the significant difference quoted above in the cases of discrepancy between BIF and Mild MR diagnoses, where IQ has a main incidence.

4. Discussion and conclusions

The standardized algorithm applied by the software supplies a diagnosis widely agreeing (for almost two third) with that based on clinical evalua-

tion. The two evaluations result, especially with respect to correspondence analysis, largely overlapping, with differences justifying the comparison between different diagnostic approaches.

The discrepancies regard always adjacent levels, particularly between Mild MR and BIF at one side, Mild and Moderate MR on the other side.

When the software suggests a diagnosis of BIF while clinical evaluation proposes Mild MR, a slight influence by the age, and a strong influence by IQ are shown. I.e., the standardized algorithm pays more attention to the younger age and to the higher IQ, while clinical evaluation tends to advantage older ages and into taking less in account the IQ, preferring to give importance to other variables, different from those emerging from the psychometric instruments.

Only in this case of discrepancy, the incidence of IQ in determining the diagnosis appears to differentiate clinical from psychometric approach. We can infer that the difference between the two modalities of evaluation is linked not to the IQ (except for the quoted discordance at the very unsteady border between BIF and Mild MR), but to other elements derived from adaptive and contextual factors.

In conclusion. the two evaluations, somewhat different although largely overlapping, show both signs of strength and weakness.

The clinical evaluation can take in to account, in a qualitative way, factors passing unnoticed unavoidably in the use of psychometric scores: e.g., developmental features, role of the context, familiar and environmental support, emotional and motivational components.

The evaluation based only on the psychometric tests scores (including - besides intelligence - also the adaptive areas of communication and socialization, as requested by the shared diagnostic criteria) has the advantage of allowing an highly standardized integration of these scores. Therefore it is undoubtedly an aid for professionals (health service employees, doctors, psychologists) with less clinical experience or for the (many) cases of diagnoses that are uncertain or divergent among different evaluators; while the formal diagnosis has to meet uniformity of criteria, considering its juridical and rehabilitative consequences, and also for prognostic purposes.

At last, we have to consider that the evaluation made by the software on the basis of the standardized algorithm is proposed as a diagnostic hypothesis, modifiable by the team on the basis of observational information and/or other diagnostic instruments. The results of other cognitive tests, as Raven's Progressive Matrices (Raven, 1949) or Leiter Scale (Leiter, 1979) where the subject's performance could be more efficient, may be taken into account. These scores may be considered alternative to the traditional IQ, also using the highly standardized conversion criteria supplied by the software itself.

References

Aa. Vv. (2007). Misure diverse di intelligenza e di adattamento nella diagnosi del grado di ritardo mentale: un tentativo di sintesi. *Ciclo Evolutivo e Disabilità / Life Span & Disability, 10* (2), 195-214

American Psychiatric Association (2000). *Diagnostic Statistic Manual, text revised (DSM-IV-TR)*, Washington: A.P.A.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis.* London: Academic Press.

Leiter, R. G. (1979). *Leiter International Performance Scale*. Chicago: Stoelting. Revised by G. H. Roid & L. J. Miller, 1997.

Shalock, R. L., & Luckasson, R. (2005). *Clinical judgment.* Washington, DC: American Association of Mental Retardation.

Luckasson, R., Borthwick-Duffy, S., Buntx, W. H. E., Coulter, D. L., Craig, E. M., Reeve, A., Schalock, R. L., Snell, M. E. Spitalnik, D. M. Spreta, S., & Tasse, M. J. (2002). *Mental Retardation: Definition, classification, and system of supports (10th ed.)*. Washington, DC: American Association of Mental Retardation.

Raven, J. C. (1949). *Guide to using the Coloured Progressive Matrices. Sets A, Ab B.* London: H. K. Lewis.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales.* Circle Pines: American Guidance Service. Tr. it. della *Forma completa – Intervista*, a cura di G. Balboni & L. Pedrabissi, Firenze: O.S., 2003.

W.H.O. (1992). *The ICD-10 Classification of mental and behavioural disorders: clinical description and diagnostic guidelines.* Geneva: World Health Organization.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition*. San Antonio: The Psychological Corporation, TX.