J Autism Dev Disord (2010) 40:334–341
DOI 10.1007/s10803-009-0877-2

ORIGINAL PAPER

Further Empirical Data on the Psychoeducational Profile-Revised (PEP-R): Reliability and Validation with the Vineland Adaptive Behavior Scales

Susanna Villa · Enrico Micheli · Laura Villa · Valentina Pastore · Alessandro Crippa · Massimo Molteni

Published online: 24 September 2009

Abstract The PEP-R (psychoeducational profile revised) is an instrument that has been used in many countries to assess abilities and formulate treatment programs for children with autism and related developmental disorders. To the end to provide further information on the PEP-R's psychometric properties, a large sample ($N = 137$) of children presenting Autistic Disorder symptoms under the age of 12 years, including low-functioning individuals, was examined. Results yielded data of interest especially in terms of: Cronbach's alpha, interrater reliability, and validation with the Vineland Adaptive Behavior Scales. These findings help complete the instrument's statistical description and augment its usefulness, not only in designing treatment programs for these individuals, but also as an instrument for verifying the efficacy of intervention.

Keywords Autism · Assessment · Psychometrics properties · Psychoeducational profile- revised (PEP-R) · Vineland Adaptive Behavior Scales

Enrico Micheli is deceased.

S. Villa (✉)
I.R.C.C.S. Eugenio Medea, Associazione La Nostra Famiglia, Via Costa Alta 37, Conegliano, TV, Italy
e-mail: susanna.villa@cn.lnf.it

E. Micheli
Laboratorio Psicoeducativo, Agordo, BL, Italy

L. Villa · V. Pastore · A. Crippa · M. Molteni
I.R.C.C.S. Eugenio Medea, Bosisio Parini, LC, Italy

Introduction

The PEP-R (psychoeducational profile revised) was developed in 1990 by Schopler et al. from an earlier version of the PEP (Schopler and Reichler 1979); it is an instrument used to assess developmental level in children with autism and other related disorders, and to design individualized training plans for them (Schopler et al. 1990).

The PEP-R has been used in many countries—as shown by its many translations in various world languages, such as Chinese (Shek et al. 2005; Ka-ting Lam and Rao 1993), Estonian (Kikas and Häidkind 2003), Dutch (Steerneman et al. 1997), and Brazilian (de Leon et al. 2004).

The PEP-R was not originally designed to determine a child's general level of intellectual functioning, nor as a psychometric instrument. Its general purpose, in fact, is to describe the typically uneven and idiosyncratic learning profiles of children with autism or related developmental disorders (Schopler et al. 1990) and therefore, to assist rehabilitation personnel in developing personalized intervention strategies for an individual child. Hence, the PEP-R's main characteristic is ease and flexibility of administration, as it was designed with these children's difficulties in mind. Indeed, is well known that using standardized tests to determine developmental level or IQ in this domain is an arduous task: a large proportion of autistic children tend to be nonverbal, have limited attention skills and poor concentration, and are easily distressed by formal assessment settings (Magiati and Howlin 2001). A directly related issue is that children who do manage to tolerate standardized test sessions—typically "mild" or "high-functioning" children who do not exhibit severe problem behaviors—are not representative of the overall population of children with autism and related developmental disorders.

For these reasons, PEP-R use has also been extended to research work, for example, to describe clinical features in case reports (Elia et al. 2000; Ramelli et al. 2008) and to validate treatment outcomes (Ozonoff and Cathcart 1998; Panerai et al. 2002, 2009; Arduino 2005), in spite of the fact that the instruments' reliability and validity to date have been somewhat limited.

The need to have valid research tools available to assess non-collaborative children with low cognitive functioning has therefore led some researchers to attempt to verify the PEP-R's psychometric properties. For example, Steerneman et al. (1997) investigated the PEP-R's interrater reliability, Cronbach's alpha, and its correlation with a Dutch non-verbal intelligence test—the *SON* 2½ – 7 (Snijders and Snijders-Oomens 1975)—in 20 children with ASD, 20 children with PDDNOS, and 20 typically developing children, (mean age 57 months.) and found encouraging results. Shek et al. (2005) closely examined the Chinese version of the PEP-R—*the CPEP-R*. Specifically, 63 children, aged 3–5 years 11 months with a diagnosis of Autistic Disorder or PDDNOS were assessed with the CPEP (Shek et al. 2005), the CARS (Schopler et al. 1988), the Merril Palmer Scale of Mental Tests (Stutsman 1948), and the Hong Kong Based Adaptive Behavior Scale (Kwok et al. 1989), which is based on the Vineland Adaptive Behavior Scales (Sparrow et al. 1984). This work yielded many helpful findings concerning interrater and test–retest reliabilities, internal consistency, and concurrent validity. Moreover, this study yielded interesting cross-cultural validity indications for the PEP-R. Another study (Delmolino 2006) investigated the possibility of using the PEP-R to evaluate young autistic children's cognitive abilities. Twenty-seven 37- to 60-month-old children with a diagnosis of Autistic Disorder or PDDNOS were assessed with the PEP-R and the Stanford Binet Intelligence Scales–fourth Edition (Thorndike et al. 1986). The results showed a PEP-R developmental quotient correlation of 0.73 with the Stanford–Binet composite IQ and verbal reasoning scores. These findings therefore allowed the author to suggest the use of PEP-R to estimate cognitive abilities in young children with these disabilities.

A closer examination of the above-described literature review, however, also reveals how many studies conducted in clinical settings are limited by a low number of participants. Moreover, the above-mentioned problem of sample representativeness remains a problem in this field, given that children capable of sitting through a directly administered comparative test session are more frequently high-functioning children.

The first goal of the present work was to provide further information on the PEP-R's reliability, especially in terms of its internal consistency and inter-rater reliability, in a large sample that also included low-functioning children. The study recruited children with a diagnosis of Autistic Disorder or PDDNOS below the age of 12 years, during a period that ranged from 2001 to 2005, at the Developmental Psychopathology Unit of the *Eugenio Medea* Scientific Institute. We opted to include all children who were assessable via PEP-R in the study—i.e., children under the age of 7 years and 7- to 12-year-old children with developmental skills below preschool-level, based on test manual indications—with the aim of obtaining evidence on the test's reliability and validity in a population presenting different levels of functioning and in which PEP-R is typically used for clinical purposes.

A second purpose of the study was to gain further knowledge on the PEP-R's concurrent validity with the Vineland Adaptive Behavior Scales. The importance of conducting a comparative study with the PEP-R Developmental Scale and the Vineland Scales lies in the fact that the latter are strongly recommended to assess Autistic Disorder (Klin et al. 2005). Moreover various studies have examined the Vineland scores of children with autism and studied the effects of age, IQ, and severity of the symptomatic triad on adaptive behavior as measured with the Vineland Scales (Cicchetti et al. 1991; Burack and Volkmar 1992; Schatz and Hamdan-Allen 1995; Carter et al. 1998).

Our own study was aimed at investigating the relation between Vineland and PEP-R scores, in two specific ways: by examining global score correlations on the two tests and by subdividing the sample based on the participants' PEP-R scores, to verify whether the test score correlations would be influenced by a greater or lower level of PEP-R-measured functioning.

A further aim was to verify whether the means of the two tests (Vineland e PEP-R) would differ in this rather large sample, given that clinical evidence frequently shows score discrepancies for the two tests in single individuals.

Method

Participants

One hundred and thirty-seven (137) children under the age of 12 years were recruited for this study. The sample derived from the clinical population that arrived at our Developmental Psychopathology Unit in the *Eugenio Medea* Scientific Institute during a period spanning from 2001 through 2005, to confirm diagnostic hypotheses and to receive treatment indications. The center is located in northern Italy and is specialized in the diagnosis and treatment of childhood disability. Many children are sent to our institute from different regions in Italy and in fact, the patients attending the center frequently represent complex cases, which are difficult for their own local social services

to manage. The children's hospitalization period lasted 15 days on average, and during this time a multidisciplinary team comprising child neuropsychiatrists, child psychologists, speech therapists, occupational therapists and special educators used a diagnostic and assessment protocol, in accordance with guidelines suggested by the Child Neurology Society and by the American Academy of Neurology 2000 (Filipek et al. 2000).

Autistic and PDD NOS participants were consecutively recruited. Diagnoses for patients in our sample were formulated at admission by child neuropsychiatrists, in accordance with the criteria described in the International Classification of Disease (ICD-10, World Health Organization 1992) and the Diagnostic and Statistical Manual of Mental Disorders (DSM IV-TR, American Psychiatric Association 2000). The diagnoses were then confirmed independently by a child psychologist through direct observation and discussion with each child's parents. Only children for whom the diagnoses concurred were included in the study. Given that the PEP-R is most appropriately used for the ages of 6 months to 7 years but can also provide helpful information for 7- to 12-year-old children who have developmental skills that do not go beyond preschool-level (Schopler et al. 1990), 7- to 12-year-old children were included only if their functional abilities were found to be below the test's maximum values.

The total sample ($N = 137$) consisted of 115 boys (83.9%) and 22 girls (16.1%), aged 25–168 months (Mean = 71.78; DS = 30.54). One hundred and twenty participants (87.6%) had received a diagnosis of Autistic Disorder, and the other 17 patients (12.4%) had been diagnosed with pervasive developmental disorder not otherwise specified (PDD-NOS).

In terms of severity of autism, the participants' mean CARS score was 40.69 (DS = 6.46); the sample's average autism level was therefore in the severe range.

Scores for thirty of the participants (28 boys and 2 girls), aged 25–118 months (mean = 68.9; DS = 23.34), were randomly selected for the inter-rater reliability analysis. This subsample's diagnoses were: Autistic Disorder for 27 participants (90%) and PDD-NOS for 3 children (10%); the mean subsample CARS score was 40.75 (DS = 4.87).

Instruments

The psychoeducational profile-revised (PEP-R) was developed by Schopler et al. 1990 as a part of their TEACCH Program (treatment and education of autistic and related communication handicapped children; Schopler 1997) to assess children with pervasive developmental disorders and other related communication disorders and to formulate individualized training plans for them. The test is most appropriately used for the ages of 6 months to 7 years; the PEP-R can also provide helpful information for 7- to 12-year-old children who have developmental skills that do not go beyond preschool-level.

The PEP-R provides information on developmental level and helps clinicians identify particular strengths and weaknesses in the following areas: imitation (16 items), perception (13 items), fine motor (16 items), gross motor (18 items), eye-hand coordination (15 items), cognitive performance (26 items), and cognitive verbal (27 items). Overall, these seven scales make up the Developmental Scale, which yields a Developmental Age measure. The PEP-R also includes a Behavioral Scale, which is used to identify the degree of behavioral abnormality and atypical behavioral characteristics of a child with autism. This scale covers four areas: relating and affect (12 items), play and interest in materials (8 items), sensory responses (12 items), and language (11 items).

The scoring system for both the scales is divided into three levels: passing, emerging, and failing for the Developmental Scale; and appropriate, mild, and severe for the Behavioral Scale. The Developmental Score is the sum of all individual item passing scores on the Developmental Scale and yields a standardized Developmental Age score.

We used an Italian version of the PEP-R, which had been developed in our institute. Correct item translation had been ensured via back-translation, and several adjustments had subsequently been made, based on email correspondence with the test's authors.

The Vineland Adaptive Behavioral Scales, Expanded Form (Sparrow et al. 1984) was employed to assess the PEP-R's concurrent validity. This instrument assesses individual autonomy and social responsibility from the first years of life through adulthood. A trained clinician or researcher administers the scales via semi-structured interview to the main caregiver. Specifically, the Vineland Adaptive Behavior Scales measure adaptive behavior in four domains: communication, daily skills, socialization, and motor skills. The Composite Scale is made up of the first three scales, and the Total Scale comprises all four scales. Each item is scored as 2, 1, or 0. It is also possible to translate single scale-, composite-, and total scores into age-equivalent scores. Our study used the Italian version of the Vineland Adaptive Behavior Scales and its normative data (Balboni et al. 2001). The lowest age-equivalent score in the Italian standardization corresponds to 18 months. Thus, raw scores were available for all of the 137 participants, but were translated into age-equivalent scores for only 72 of the children.

Procedure

A multidisciplinary team made up of child neuropsychiatrists, child psychologists, speech therapists, occupational

therapists, and special educators used a diagnostic and assessment protocol, in accordance with guidelines suggested by the Child Neurology Society and by the American Academy of Neurology (Filipek et al. 2000), to individually assess children with autism or PDD during their (on-average) 15-day stay at the Scientific Institute. The PEP-R, the CARS, and the VABS were included in this protocol. The PEP-R was administered to all the children below 7 years of age, and to the 7- to -12-year-old children presenting developmental skills below preschool-level. Specifically, only the 7- to 12-year-old children whose imitation, cognitive performance, and cognitive verbal subscale scores were found to be below each subscale's maximum limit—16, 26, and 27, respectively—were included. The PEP-R and Vineland scores are shown in Table 1.

All the children in the sample were administered the PEP-R in an assessment room by an examiner with many years of experience in test administration procedures and in working with children with autism. The room was fitted with a one-way mirror behind which a psychologist observed and wrote down each child's responses, during fixed system video-recording. The participants' parents also watched the proceedings together with the psychologist and were asked to mention any information that might help clarify the child's behavior during the test. The procedure was therefore based on the idea that clinical assessment in autism should include many information sources (Klin et al. 1997).

To examine inter-rater reliability, Excel's "Randomize" function was used to randomly select 30 participants, and their video-recordings were independently evaluated by two psychologists, who had been trained in PEP-R scoring procedures. These two evaluators, however, were blind to the participants' diagnoses and had no contact with the children's families or clinicians. The aim of this procedure was to assess inter-rater agreement based only on test-manual information and on directly observable responses.

Results

Reliability

Cronbach's (1951) alpha coefficient was used to estimate internal consistency for the various PEP-R domains; the resulting alpha values are shown in Table 2.

Cronbach's alphas were excellent for all the Developmental Scales (with the exception of Perception $-\alpha = .84$) and for the Behavioral Scale of language, ranging from .90 to .99. Alpha values for the other Behavioral Scales and for the Developmental Scale of perception ranged from .82 to .85 and can therefore be considered good, based on Cicchetti's (1994) and Cicchetti and Sparrow's (1981) criteria.

The intraclass correlation coefficient was used as an inter-rater agreement index for the subgroup randomly selected for the inter-rater reliability analysis. No

Table 1 Descriptive information—participant PEP-R scale and Vineland Adaptive Behavior Scale (VABS) scores

	Total sample $N = 137$	Age 0–3 $n = 7$	Age 3–6 $n = 77$	Age 6–9 $n = 33$	Age 9–12 $n = 20$
PEP-R developmental scale					
Imitation	4.93 (4.52)	1.14 (1.34)	4.27 (4.24)	6.58 (4.80)	6.05 (4.67)
Perception	8.69 (3.20)	4.43 (1.81)	8.44 (2.96)	9.76 (3.39)	9.35 (2.89)
Fine motor	9.70 (3.57)	4.71 (2.92)	9.10 (3.33)	11.12 (3.22)	11.40 (3.12)
Gross motor	13.08 (4.24)	6.43 (2.30)	12.44 (4.02)	14.73 (3.74)	15.15 (3.25)
Eye-hand coordination	6.33 (3.77)	2.43 (1.81)	5.57 (3.11)	7.58 (3.89)	8.55 (4.52)
Cognitive performance	6.93 (6.65)	0.71 (0.49)	5.78 (5.89)	9.61 (7.35)	9.15 (6.98)
Cognitive verbal	3.76 (5.79)	0.71 (0.49)	2.78 (4.41)	5.88 (7.86)	5.10 (6.40)
Developmental score	53.44 (27.93)	20.57 (8.89)	48.40 (24.61)	65.27 (30.40)	64.80 (26.43)
Developmental age (months)	24.62 (11.68)	13.00 (4.20)	22.65 (9.81)	29.39 (14.04)	28.40 (11.57)
VABS scale raw scores					
Communication	56.20 (46.34)	16.71 (7.78)	50.87 (39.71)	72.97 (56.93)	62.85 (48.55)
Daily skills	79.73 (40.55)	23.29 (12.13)	71.51 (33.27)	94.00 (42.60)	107.60 (40.26)
Socialization	41.99 (19.96)	19.86 (9.47)	41.51 (18.17)	48.58 (31.94)	40.70 (20.84)
Motor skills	95.94 (22.70)	53.43 (14.00)	95.10 (19.10)	103.15 (23.01)	102.15 (21.83)
Composite	178.88 (99.50)	58.43 (25.90)	164.56 (84.51)	218.30 (114.25)	211.15 (100.60)
Total	275.63 (127.43)	110.43 (36.27)	261.61 (118.23)	320.55 (134.55)	313.30 (116.42)

Table 2 PEP-R internal consistency and inter-rater reliability

	Alpha ($N = 137$)	Inter-rater reliability ($N = 30$)
Developmental Scale		
Imitation	.94	.95
Perception	.84	.84
Fine motor	.91	.97
Gross motor	.90	.95
Eye-hand coordination	.93	.97
Cognitive performance	.96	.98
Cognitive verbal	.97	.88
Developmental Score	.99	.98
Behavioral Scale		
Relating and affect	.85	.65
Play and interest in materials	.83	.56
Sensory responses	.82	.74
Language	.99	.87

significant differences between the randomly selected subgroup and the total sample concerning age ($t = .636$; $p = .526$), gender, ($\chi^2 = 2.07$; $p = .15$), diagnosis ($\chi^2 = 0.791$; $p = 0.673$) CARS score ($t = -.025$; $p = .98$), VABS score ($t = 0.024$; $p = .98$), or PEP-R score ($t = .024$; $p = .60$) were yielded.

Table 2 also shows the inter-rater reliability results. The intraclass correlation coefficients values for the Developmental Domain ranged from .84 to .98 and can be considered excellent, based on Cicchetti's (1994) and Cicchetti and Sparrow's (1981) criteria. The values were lower for the Behavioral Domain, with coefficients ranging from .56 to .87.

Validity

The Pearson correlation coefficient between the PEP-R Developmental Scale (Developmental Score and Developmental Age) and Vineland Domain Scores was used to examine the PEP-R's concurrent validity (Table 3).

As described above, the lowest Vineland Scale Domain age-equivalent score on the Italian standardization corresponds to 18 months, and 72 of our 137 participants reached this minimum score in all domains. We therefore considered raw scores, according to a previously experimented procedure that compared the PEP (1979-version) with raw scores on the handicap behavior and skills schedule, an adaptive behavior measure derived from the Vineland Scales (van Berckelaer-Onnes and van Duijn 1993).

As shown in Table 3, both the PEP-R Developmental Score and Developmental Age score correlated significantly

Table 3 Pearson correlation coefficients between PEP-R (Developmental Score and Developmental Age) and Vineland Adaptive Behavior Scale Domain Scores (VABS) ($n = 137$)

Domains Scores of the Vineland Adaptive Behaviour Scales (VABS)	PEP-R Developmental Score	PEP-R Developmental Age
Communication	.83*	.84*
Daily skills	.81*	.75*
Socialization	.78*	.75*
Motor skills	.82*	.76*
Composite	.87*	.85*
Total	.85*	.82*

* $p < .001$

($p < .001$) with the Vineland's total, composite, and domain scores.

The correlation size between each Vineland Domain score and PEP-R Developmental Score ranged from 0.78 (Vineland socialization and PEP-R Developmental Score) to 0.87 (Vineland composite and PEP-R Developmental Score). The Vineland and PEP-R Developmental Age score correlations ranged from 0.75 (daily living skills and socialization) to 0.85 (Vineland composite scores). These values show a very large overall correlation, based on the criteria used by McCarthy et al. (1991), who expanded on Cohen's (1988) effect size criteria.

The total sample used to examine concurrent validity was then split at the median PEP-R Developmental Age value (21 months), and two participant groups were determined thereby: a "high functioning" group (Developmental Age > median) and a "low functioning" group (Developmental Age < median). Correlation coefficients between the PEP-R (Developmental Score and Developmental Age) and Vineland Domain Scores were calculated once more for each group. Chronological age was used as a control for partial correlations (Table 4). The resulting correlation coefficients were generally higher for the "low functioning" group although the between-group difference was not significant.

Lastly, a *T*-Test between mean the PEP-R Developmental Age ($m = 31.41$, SD = 11.68) Scores and Vineland Total Domain Scores ($m = 30.30$; SD = 8.28) was calculated for a group of 72 children presenting a Vineland Total Domain Score of > 18 months. The difference was not significant ($T = -0.662$; $p = 0.254$), but we found a high degree of within-group variability: more than 83.3% of the participants showed a score discrepancy on the two tests of at least 12 developmental months. Half of these had higher Vineland scores, and the other half had higher PEP-R scores, even though the entire sample's score means showed no significant differences.

Table 4 Partial correlations for total sample split at PEP-R Developmental Age median value, with chronological age as a control ($n = 137$)

Vineland Adaptive Behavior Scale Domain (VABS)	PEP-R Developmental Score	PEP-R Developmental Age
Low-functioning group		
Communication	.52*	.48*
Daily skills	.71*	.67*
Socialization	.66*	.64*
Motor skills	.74*	.72*
Composite	.71*	.67*
Total	.77*	.73*
High-functioning group		
Communication	.75*	.73*
Daily skills	.51*	.47*
Socialization	.62*	.59*
Motor skills	.58*	.54*
Composite	.72*	.69*
Total	.67*	.64*

* $p < .001$

Discussion

This study's findings appear difficult to disentangle because our participants were not preliminarily selected by singling out different groups of children presenting homogeneous traits.

Yet, the first goal of this study was to generate further data on the PEP-R's psychometric properties, so as to complete the statistical description of this instrument for its typical clinical use—i.e., with high or low functioning children under the age of 7 years or with low-functioning 7- to 12-year-old children. It is important to note that (as mentioned in the "Introduction") the PEP-R has also been used as an instrument for evaluating treatment outcomes in this type of population (Ozonoff and Cathcart 1998; Panerai et al. 2002, 2009; Arduino 2005).

We therefore required a rather composite sample and selected one that included younger children with different levels of functioning as well as low-functioning older children (aged 7–12 years) with ability levels below those measured by the PEP-R. Indeed, we believe that the choice to not conduct preliminary participant selection represents the study's originality: the sample included all the children at our institute who were consecutively assessed via PEP-R during the period spanning from 2001 through 2005: It is therefore [highly] likely that this sample presented the characteristics of the population that is assessable with this instrument, and indeed, we believe it is useful to evaluate the psychometric characteristics of an instrument starting from the population upon which it is usually used. Moreover, researchers analyzing the psychometric characteristics of the PEP-R by examining samples of children under the age of 7 years with high level of functioning, have also highlighted the need in future research to focus on validity and reliability analyses in samples presenting wider ranges of functioning and age (Steerneman et al. 1997; Muris et al. 1997; Shek et al. 2005).

With respect to reliability, several satisfactory and high Cronbach's alpha values were yielded, similarly to other studies conducted on smaller samples with a restricted range of intellectual level (Shek et al. 2005; Steerneman et al. 1997).

Interrater reliability scores for the Behavioral Domain were lower than they were for the Developmental Domain, which yielded an excellent effect size, based on Cicchetti's (1994) and Cicchetti and Sparrow's (1981) criteria. Other studies (Shek et al. 2005; Muris et al. 1997) have also shown lower Behavioral Domain interrater reliability than observed for the Developmental Domain, but the difference in the present study was of a higher magnitude. A new version of the test attempting to breach this gap has been recently published (the PEP-3, Schopler et al. 2005). With respect to the previous version, it presents a more detailed Behavioral Scale, in terms of both administration and scoring procedures, and easier interpretation based on more clearly defined criteria.

Secondly, the Vineland Scale score comparison yielded some very interesting results. Based on findings in the literature concerning comparisons between previous or current versions of the PEP-R and Vineland Scales or with other instruments used for adaptive behavior assessment (Schopler et al. 1990; Van Berckelaer-Onnes and van Duijn 1993; Shek et al. 2005), we expected to observe significant between-score correlations. This hypothesis was confirmed, in the sense that the two tests yielded good correlations, which were slightly higher in the total domain, however, for lower general level of functioning than they were for the scores of higher level functioning children. It is interesting to note that, in the present study, both the PEP-R and the Vineland yielded similar group mean age-equivalent scores, but a high degree of intraindividual variability was observed. Although, as described above, the means comparison yielded essentially overlapping mean age-equivalent scores for the two measures, an examination of each child's scores showed a difference of >12 months between the age-equivalent scores on the two tests in 83.3% of the instances. This finding has strong implications for research and/or clinical practice: although either test may be used with large participant samples to estimate the sample's average performance, it must not be forgotten that the two measures are not interchangeable at the single participant level. In fact, children undergoing PEP-R assessment perform in a structured situation, but Vineland

assessment requires that parents rate their children in everyday life situations. As stated by Van Berckelaer-Onnes and Van Dujin (1993), "using both scales clinicians and parents share a serious responsibility to each other". Information obtained both through parents and via objective assessment is needed to design individual treatment programs and to evaluate their efficacy (e.g., with a single participant experimental design). This general affirmation concerning the importance of tapping multiple sources during the assessment process (Klin et al. 1997) has even more concrete implications when intra-individual differences in scores obtained via different instruments emerge.

Lastly, in our study these intraindividual differences went in opposite directions: half of the 83% children showing a >12 month age-equivalent score difference on the two measures showed this difference on the Vineland Scales and the other half showed it on the PEP-R.

This situation, however, is well-known among people who work daily in clinical practice. The generally accepted reason is that some children perform best in well-structured settings (and therefore tend to score higher on the PEP-R), but other children do better in their daily life contexts (especially when familiar people are available to provide various types of environmental facilitation) and therefore usually score higher on the Vineland Scales. Hence, this finding suggests that, whereas only one test might suffice in determining age-equivalent scores when conducting research on large groups, different instruments are required to assess individual functioning and to design individual treatment programs.

We note that our study has two important limits. The first is a direct consequence of collecting data from a sample of children without having conducted a priori selection: This choice presented the methodological weakness of not having IQ scores available for analysis, though, it was not possible to obtain them because many of the participants could not be assessed via standardized IQ instruments. The second limitation derives from the period in which the study was conducted—2001 through 2005—a period in which the authorized versions of the Autism Diagnostic Observation Scale (ADOS, Lord et al. 1989) and of the standardized interview Autism Diagnostic Interview-Revised (Lord et al. 1994) were not yet available in Italy (they were published in Italy only in 2005). Being able to use these instruments would have allowed for gold standard diagnostic confirmation.

Despite these limits, however, these further empirical data on the validity and reliability of the PEP-R can be considered valuable because the instrument has enjoyed wide-scale use, not only in assessment and treatment planning, but also for research purposes, especially in studies aimed at verifying the efficacy of intervention. The PEP-R has been recently updated and replaced by PEP-3 (Schopler et al. 2005). Yet, the PEP-R has been used for so many years in the field of autism, in both its clinical and research domains, that we believe that knowledge of its statistical properties can still serve a highly useful purpose for longitudinal research using this instrument.

In any event, our future research aims include a comparison of the PEP-R and PEP3—i.e., by comparing the data of both instruments with the Vineland and/or other measures—so as to evaluate the potential comparability of the data gathered via the two instruments.

Lastly, and perhaps most importantly, our study highlights the efficacy of collecting data in clinical settings in order to validate assessment instruments with characteristics of flexibility in assessing children with developmental difficulties. Indeed, it is crucial to include children who cannot be assessed with standardized tests—due to their non-compliance and idiosyncratic developmental and behavioral characteristics—in future research. It is hoped that our study will serve to stimulate reflection on the need to have empirically reliable and valid clinical instruments that can concurrently keep low-functioning children from remaining "under the radar".

Acknowledgments The authors would like to thank the colleagues of the Developmental Psychopathology Unit of the *Eugenio Medea* Scientific Institute for their cooperation in the study, especially Elisa Mani, Cristina Motta, Giuseppe Aceti, Barbara Urbani, Maura Mariani, Rosaria Fontana, Stefano Buson, Antonella Terraneo, Moira Mambretti, Silvia Saporiti, Piera Carganico and Maria Nobile.

References

American Academy of Neurology. (2000). Practice parameter: Screening and diagnosis of autistic spectrum disorders. *Neurology, 55*(4), 468–479.

American Psychiatric Association. (2000). DSM IV-TR *diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: American Psychiatric Association.

Arduino, M. (2005). Valutazione degli esiti: uno studio attraverso il PEP-R. *Autismo e disturbi dello sviluppo, 3*(1), 99–126.

Balboni, G., Pedrabissi, L., Molteni, M., & Villa, S. (2001). Discriminant validity of the Vineland Scales: Score profiles of individuals with mental retardation and a specific disorder. *American Journal of Mental Retardation, 106*(2), 162–172.

Burack, J. A., & Volkmar, F. R. (1992). Development of low- and high-functioning autistic children. *Journal of Child Psychology and Psychiatry, 33*, 607–616.

Carter, A. S., Volkmar, F. R., Sparrow, S. S., Wang, J. J., Lord, C., Dawson, G., et al. (1998). The Vineland Adaptive Behavior Scales: Supplementary norms for individuals with autism. *Journal of Autism and Developmental Disorders, 28*(4), 287–302.

Cicchetti, D. V. (1994). Guidelines, criteria and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290.

Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: Assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127–137.

Cicchetti, D. V., Sparrow, S. S., & Rourke, B. P. (1991). Adaptive behavior profiles of psychologically disturbed and developmentally disables children. In J. L. Matson & J. Mulich (Eds.), *Handbook of mental retardation (Chap. 15* (pp. 222–239). New York: Pergamon.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

De Leon, V., Bosa, C., Hugo, C., & Hutz, C. S. (2004). Psychometric properties of the psychoeducational profile revised: PEP-R. *Avaliação Psicologica, 3*(1), 39–52.

Delmolino, L. M. (2006). Brief report: Use of DQ for estimating cognitive ability in young children with autism. *Journal of Autism and Developmental Disorders, 36*(7), 959–963.

Elia, M., Ferri, R., Musumeci, S. A., Panerai, S., Bottitta, M., & Scuderi, C. (2000). Clinical correlates of brain morphometric features of subjects with low-functioning autistic disorder. *Journal of Child Neurology, 15*(8), 504–508.

Filipek, P. A., Accardo, P. J., Ashwal, S., Baranek, G. T., Cook, E. H., Dawson, G., et al. (2000). Practice parameter: Screening and diagnosis of autism. *Neurology, 55*, 468–479.

Ka-ting Lam, M., & Rao, N. (1993). Developing a Chinese version of the psychoeducational profile (CPEP) to assess autistic children in Hong Kong. *Journal of Autism and Developmental Disorders, 23*(2), 273–279.

Kikas, E., & Häidkind, P. (2003). Letter to the editor: Developing an Estonian version of the psychoeducational profile revised (PEP-R). *Journal of Autism and Developmental Disorders, 33*(2), 217.

Klin, A., Carter, A., Volkmar, F. R., Cohen, D. J., Marans, W. D., & Sparrow, S. (1997). Assessment issues in children with autism. In D. Cohen & F. Wolkmar (Eds.), *Handbook of autism and pervasive developmental disorders* (2nd ed., pp. 411–415). NY: John Wiley & Sons.

Klin, A., Saulnier, C., Tsatsanis, K., & Volkmar, F. (2005). Clinical evaluation in autism spectrum disordered: Psychological assessment within a transdisciplinary framework. In F. R. Volkmar, R. Paul, A. Klein, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (3rd ed.). NY: John Wiley & Sons.

Kwok, J., Shek, D. T. L., Tse, J., & Chan, S. (1989). *Hong Kong based adaptive behavior scale*. Hong Kong: Department of Applied Social Studies, City Polytechnic of Hong Kong.

Lord, C., Rutter, M., Goode, S., et al. (1989). Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders, 19*(2), 185–212.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A recise version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659–685.

Magiati, I., & Howlin, P. (2001). Monitoring the progress of preschool children with autism enrolled in early intervention programmes: Problems in cognitive assessment. *Autism, 5*(4), 399–406.

McCarthy, P. L., Cicchetti, D. V., Sznajderman, S. D., Forsyth, B. C., Baron, M. A., Fink, H. D., et al. (1991). Demographic, clinical and psychosocial predictors of the reliability of mothers' clinical judgments. *Pediatrics, 88*, 1041–1046.

Muris, P., Steerneman, P., & Ratering, E. (1997). Brief report: Interrater reliability of the psychoeducational profile (PEP). *Journal of Autism and Developmental Disorders, 27*(5), 621–626.

Ozonoff, S., & Cathcart, K. (1998). Effectiveness of a home program intervention for young children with autism. *Journal of Autism and Developmental Disorders, 28*(1), 25–32.

Panerai, S., Ferrante, L., & Zingale, M. (2002). Benefits of the treatment and education of autistic and communication handicapped children (TEACCH) programme as compared with a non-specific approach. *Journal of Intellectual Disability Research, 46*(4), 318–327.

Panerai, S., Zingale, M., Trubia, G., Finocchiaro, M., Zuccarello, R., Ferri, R., & Elia, M. (2009). Special education versus inclusive education: The role of the TEACCH Program. *Journal of Autism and Developmental Disorders*, *Feb 10* (Epub ahead of print).

Ramelli, G. P., Silacci, C., Ferrarini, A., Cattaneo, C., Visconti, P., & Pescia, G. (2008). Microduplication 22q11.2 in a child with autism spectrum disorder: Clinical and genetic study. *Developmental Medicine & Child Neurology, 50*(12), 953–955.

Schatz, J., & Hamdan-Allen, G. (1995). Effects of age and IQ on adaptive behavior domains for children with autism. *Journal of Autism and Developmental Disorders, 25*(1), 51–60.

Schopler, E. (1997). Implementation of TEACCH philosophy. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive developmental disorders* (2nd ed., pp. 767–795). NY: John Wiley & Sons.

Schopler, E., Lansing, M. D., Reichler, R. J., & Marcus, L. M. (2005). *PEP-3 psychoeducational profile* (3rd ed.). Austin: Pro-Ed.

Schopler, E., & Reichler, R. J. (1979). *Individual assessment and treatment for autistic and developmental disabled children: Vol. 1. Psychoeducational profile*. Baltimore: University Park Press.

Schopler, E., Reichler, R. J., Bashford, A., Lansing, M. D., & Marcus, L. M. (1990). *Individual assessment and treatment for autistic and developmental disabled children: Vol. 1. Psychoeducational profile revised (PEP-R)*. Austin: Pro-Ed.

Schopler, E., Reichler, R. J., & Rochen Renner, B. (1988). *The childhood autism rating scale (CARS)*. Los Angeles: Western Psychological Services.

Shek, D. T. L., Tsang, S. K. M., Lam, L. L., Tang, F. L. Y., & Cheung, P. M. P. (2005). Psychometric properties of the Chinese version of the psycho-educational profile-revised (CPEP-R). *Journal of Autism and Developmental Disorders, 35*(1), 37–44.

Snijders, J. T., & Snijders-Oomens, N. (1975). *Snijders-Oomens nietverbale intelligentietest (Snijders—Oomens non verbal intelligence test) 2½ − 7*. Gronigen: Woltesr-Noordhoff.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behaviour Scales*. Circle Pines, MN: American Guidance Service. Italian Edition by G. Balboni & L. Pedrabissi (2003). Florence: O.S. Edition.

Steerneman, P., Muris, P., Merckerlbach, H., & Willems, H. (1997). Assessment of development and abnormal behavior in children with pervasive disorders. Evidence for the reliability and validity of the revised psychoeducational profile. *Journal of Autism and Developmental Disorders, 27*(2), 177–185.

Stutsman, R. (1948). *Merrill-palmer scale of mental tests*. Los Angeles, CA: Western Psychological Services.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford Binet intelligence scale*. Itasca, IL: Riverside Publishing.

Van Berckelaer-Onnes, I., & van Duijn, G. (1993). A comparison between the handicaps behaviour and skills schedule and the psychoeducayional profile. *Journal of Autism and Developmental Disorders, 23*(2), 263–272.

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