Intelligence 38 (2010) 462–470

Contents lists available at ScienceDirect

Intelligence

The mean Southern Italian children IQ is not particularly low: A reply to R. Lynn (2010)

Cesare Cornoldi [a], Carmen Belacchi [b], David Giofrè [a,*], Angela Martini [c], Patrizio Tressoldi [a]

[a] Department of General Psychology, University of Padua, Italy
[b] Institute of Psychology, University of Urbino, Italy
[c] INVALSI, Frascati-Roma, Italy

ARTICLE INFO

Article history:
Received 23 March 2010
Received in revised form 27 May 2010
Accepted 6 June 2010
Available online 10 July 2010

Keywords:
IQ
Italy
Regional differences
PISA

ABSTRACT

Working with data from the PISA study (OECD, 2007), Lynn (2010) has argued that individuals from South Italy average an IQ approximately 10 points lower than individuals from North Italy, and has gone on to put forward a series of conclusions on the relationship between average IQ, latitude, average stature, income, etc. The present paper criticizes these conclusions and the robustness of the data from which Lynn (2010) derived the IQ scores. In particular, on the basis of recent Italian studies and our databank, we observe that : 1) school measures should be used for deriving IQ indices only in cases where contextual variables are not crucial: there is evidence that partialling out the role of contextual variables may lead to reduction or even elimination of PISA differences; in particular, schooling effects are shown through different sets of data obtained for younger grades; 2) in the case of South Italy, the PISA data may have exaggerated the differences, since data obtained with tasks similar to the PISA tasks (MT-advanced) show smaller differences; 3) national official data, obtained by INVALSI (2009a) on large numbers of primary school children, support these conclusions, suggesting that schooling may have a critical role; 4) purer measures of IQ obtained during the standardisation of Raven's Progressive Coloured Matrices also show no significant differences in IQ between children from South and North Italy.

The situation of South Italy, marked by its chequered history and its social, economic and cultural differences, is well-known to Italians but equally to researchers abroad who have studied Italy's history and development (e.g., A'Hearn, 1998). The paper by Lynn (2010) concerning North–South differences in IQ in our view displays an over-reliance on certain available data and misses some critical points about the Italian situation, in so doing incurring the risk of not only giving a not realistic view of South Italy, but also confusing the debate on heritability of human intelligence. Research has cleared up the various recurring topics linked to regional differences and racial differences in intelligence, and in the relationship between these and other variables such as geographical location, income, somatic features (here stature), health factors (here infant mortality) and literacy. In particular, Lynn (2010) started out from learning scores in reading, mathematics and science gathered under the Programme for International Student Assessment (PISA) study (Organisation for Economic Co-operation, & Development [OECD], 2007), which indicated a differential between students from North and South Italy of about 100 points, according to PISA standardisation, which sets the average at 500 and standard deviation at 100; from these data he derived an IQ index that highly correlated with a series of geographic, demographic, and social variables. Lynn (2010) summarised his conclusions under 10 points, all following from the assumption that there are innate differences in IQ between Italians from North and South Italy, and that these differences had a causal effect on the other variables.

This approach is represented by well-known stances (e.g., Herrstein & Murray, 1994; Jencks, 1972; Jensen, 1985), which

* Corresponding author. Department of General Psychology, University of Padua, via Venezia, 8, 35131, Padova, Italy.
E-mail address: david.giofre@virgilio.it (D. Giofrè).

0160-2896/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2010.06.003

have been the subject of sometimes fierce debate (see Howe, 1997). Criticisms of Lynn's work have pivoted on a wide variety of approaches, but argument has for the main part had an ideological or theoretical basis rather than empirical. Through our analysis we examine the assertions of Lynn (2010) starting out from the empirical elements he used, in particular the achievement scores and the IQs of Italian children. We are able to do this using data we have gathered directly or from collaborating agencies, and which—although partly accessible through public documents—were not used by Lynn (2010). More specifically, the data we present here are based on re-analysis of the PISA data; the relatively small data sets (MT-Advanced, Grade 9–10 tests in math and reading) from our tasks paralleling the PISA tasks (Cornoldi, Friso, & Pra Baldi, 2010); the very large INVALSI (Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione) data set based on samples representative of the populations of South and North Italy (sampled and tested according to the same methodology used by PISA); and our data sets collected during the new Italian standardisation of Coloured Progressive Matrices (Belacchi, Scalisi, Cannoni, & Cornoldi, 2008).

1. Factors affecting Pisa data

As might be expected, the difficult situation of the schools in South Italy highlighted by the PISA study has been subject of considerable contention, throughout the world as well as in Italy itself. Focus has been less on possible inferences regarding intellectual function, and more on the school achievements of Italian children, sometimes with a reference to a not well specified condition of underachievement, sometimes with a more specific reference to the clinical category of learning disability (LD). For example, on the basis on the PISA data and the standards typically used to diagnose a reading or a mathematical learning disability, about one third of children in Italy's South might have been given a diagnosis of LD, in clear contrast with the hypothesis that percentages of LD are lower than 10% and comparable in all countries worldwide (Jimenez & Garcia de la Cadena, 2007; Zorman, Lequette, & Pouget, 2004), and that in Italy, at least when it comes to reading, numbers would actually be lower than in other countries as a result of the language's transparency—i.e., perfect correspondence between how a word is written and how it is pronounced (Ziegler & Goswami, 2005). In fact, Lindgren, De Renzi, and Richman (1985) reported a lower number of reading disabled children in Italy (3.6%) than in the USA (4.5%). The general view of Italian figures considering the PISA data has been that the differences derive from the effects of sociocultural and economic factors. There is a large-scale literature showing that sociocultural and economic factors and other comparable factors have also genetic components, for instance educational performance and IQ scores are strongly genetically linked (Lynn, 2010). Therefore, it cannot be excluded that, when one corrects for family and background variables indeed also removes a sizable part of the genetic differences. However, although sociocultural and economic factors could be related with intrinsic hereditary characteristics of Italian children, they might also be due to more contingent contextual factors and therefore the analysis of their effects appear important for the present debate. For example, Bratti, Checchi, and Filippin (2007) investigated the existence and dimension of territorial differences in the maths skills of Italian students. Their analysis benefited from the data set that merges the 2003 wave of the PISA data (OECD, 2003) with territorial data collected from several statistical sources and with administrative school data from the Italian Ministry of Education. In addition to the standard gradient represented by parental education and occupation (PISA Economic, Social and Cultural Status index), they considered three different groups of educational input: individual characteristics (mainly family background), school types and available resources, and territorial features related to labour market, cultural resources and aspirations. In particular, among the local factors measured at Province level, they found a substantial impact of buildings maintenance and likelihood of employment. Student sorting across school types was also found to play a relevant role (in the North student sorting takes place according to ability and is based on school tracking and repetition: a less talented student is directed towards technical/vocational schools and/or held back one or more years; in the South students are less sorted among tracks). When accounting for territorial differences, Bratti et al. (2007; see also Santello, 2009) found that most of the North–South divide (75%) is accounted for by differences in school financial resources, teacher tenures, family background, while other contextual variables account for the remaining fraction. In conclusion, correcting for family and context background removes a sizable part of the postulated genetic differences between regional groups.

1.1. MT-Advanced data

MT-Advanced tasks are achievement tests designed for assessing 9th- and 10th-graders in their learning of Reading and Mathematics. The early version of the tasks (Cornoldi, Pra Baldi, & Rizzo, 1991) involved just a large battery of reading comprehension tasks. In 2005 it was decided to proceed with a new administration of some of the reading comprehension tasks and to add a series of maths tasks developed on the basis of the PISA tasks. In particular, the new battery (see Cornoldi et al., 2010) included the following tasks: Reading Comprehension, Arithmetic, Algebra, Geometry and Measures, Reading Decoding, Arithmetic Problem-solving, Mental Calculation, and Arithmetical Facts. For 9th- and 10th-graders, the Manual gives Cronbach's alpha as .71 and .84 for Reading Comprehension tasks, .79 and .71 for Arithmetic, .72 and .57 for Geometry and Measures, .79 for Arithmetic Problem-solving and .78 for Algebra, respectively. The Manual also gives test–retest reliability for the procedures adopted for measuring Reading Decoding Speed, .97, Reading Decoding Errors, .86, Mental Calculation Time, .75 and Arithmetical Facts, .78 (the range of reliability PISA varies from .43 to .93 [see tables 12.2, 12.6, and 12.7; OECD, 2009]).

The main goal of developing the new test battery was to produce a set of tasks suitable for comparing the performance of individual students or classes at a particular grade against typical Italian standards. In order to obtain large representative samples, applications for funding were made but without success, confirming the modest interest of Italian research agencies in basic research on school achievement. We therefore had to seek the help of collaborators and teachers as volunteers to administer the tasks in schools where access was easy. For these reasons the data concern relatively small groups of subjects and—although different areas, sociocultural levels and types of school were included—the samples are not perfectly representative of the populations concerned.

For the present analysis, in order to have a sufficiently large number of subjects, we also included subjects of Central Italy and split the overall sample into two subsamples, based on the classical separation between Italians south and north of Rome (a similar split was made by the Italian Government in the past when it created special policies for Southern Italy (see Felice, 2007), actioned through the 'Cassa del Mezzogiorno' (public funding agency for Italy's South). The most represented Regions for the North were Lombardy, Veneto, Piedmont and Emilia and, for the South, Lazio, Apulia and Sicily. Note that data were collected for grades rather than ages, but clearly ages represented for the two groups showed no differences.

The description of the test administration and the data themselves provide an overview of possible scenarios during assessment procedures and of different patterns of performance for students of North and South Italy. In particular our data take account of the observation frequently reported and also noted by our test administrators, that many students especially in the South were neither used to doing tasks such as those in the PISA study, nor motivated to doing them. For example, Quintano (2007) analysed missing PISA data, finding that over 20% of Southern Italy students had completed the questionnaire with over 10% of missing data, whereas in North Italy this occurs for only for 6–7% of students. We therefore decided to overcome the problem of low student involvement (at least partially) in two ways: (1) by eliminating protocols that reflected total disengagement in the task by producing a performance at chance level, a result which could be due to either random responses or substantially incomplete questionnaires; (2) by also including—when the testing was carried out by a member of our group rather than a teacher—individual testing, where the student had necessarily to collaborate with the experimenter. Within the project, individual testing mainly addressed basic decoding and computation skills, which—although less loaded on g factor—still represent important cognitive functions within intellectual abilities (Jordan, Hanich, & Kaplan, 2003; Lennon, 1950; Stanovich, 1991).

In conclusion, the study tested the following learning competencies within the classroom:

- Reading Comprehension, based on two passages and 20 multiple-choice questions.
- Arithmetic, based on 12 multiple-choice questions for 9th-grade and 10 for 10th-grade.
- Algebra (only for 10th-grade), based on 11 multiple-choice questions, was also considered.
- Geometry and Measures, based on 15 multiple-choice questions for 9th-grade and 10 for 10th-grade.
- Arithmetic Problem-solving (only for 9th-grade), based on 10 open questions, was included.

The following competencies were also tested in individual sessions:

- Reading Decoding Accuracy (number of errors in reading a 1123-word passage for 9th-grade and 1287-word passage for 10th-grade).
- Reading Decoding Speed (mean number of syllables read in 1 s during the reading decoding task).
- Mental Calculation Test (MC accuracy), i.e., number of correct responses (max = 8) given for calculations representing the four basic operations (two per operation).
- Mental Calculation Time (MC time), i.e., time required to do the calculations.
- Arithmetical Facts, i.e., the immediate correct responses given by subjects for 27 calculations where response should be known without calculation (e.g., multiplication facts).

Table 1 shows the mean scores obtained. For the group, testing scores reflect valid protocols. Invalid protocols were in fact only obtained during the group administration (nearly all students collaborated during individual testing) and represented a sizable group, varying according to the part of Italy represented (for example we found 63 invalid protocols out of the 817 arithmetic, algebra and geometry protocols of North Italy 9th-graders, whereas for the same tasks the invalid protocols of South Italy students were approximately double, i.e., 124 out of 728). Also students who were not born in Italy were excluded from the analysis as their performance could be affected by a poor knowledge of Italian. Table 1 also presents the standard error mean, the pooled SD, the Cohen's d, i.e. the effect size obtained by calculating the differences between North and South and dividing by pooled SDs, and the Student's t-value obtained by comparing the two groups (alpha was set at .01 in view of the high number of comparisons and large group sizes).

Table 1 gives an interesting overview of the actual differences found between Northern and Southern Italian adolescents in reading and mathematics. First we consider the tasks more directly paralleling the PISA tasks. In reading comprehension the differences in Cohen's d (0.31 and 0.43 for 9th- and 10th-graders, respectively) are clear though slightly smaller than in PISA. A similar pattern is also found for the Arithmetic scores (0.27 to 0.43), confirming but also weakening the pattern observed during the PISA 2006 study. Surprising and provocative is the pattern observed during the individual administrations; in fact, readers should avoid being misled by the typically higher values for North Italian children, since the reading decoding value actually refers to the average number of errors made, and in maths higher times obviously reflect slower performance.

Notice that some differences in group testing could have been here underestimated and, if corrected for the tests reliability (which were not particularly high), could be closer to the differences reported in the PISA study. However the disattenuation correction due to fallible measurement may produce artifacts of overcorrection (Lihshing, 2010) and, in particular, could paradoxically emphasize also the differences we found in favour of Southern Italy children. In fact some differences clearly go in a completely different direction: a closer look at the results from individual testing now shows that for reading, the children of North Italy are brighter (the differences in Cohen's d, with respect to Southern Italy children, are 0.36 and 0.30 for 9th- and 10th-graders, respectively) but also less accurate, making approximately 50% more errors (values are impressively high, at 0.66 and 0.84, respectively). In the case of mathematics, only one difference between North and South is significant (mental calculation accuracy for 9th-graders) a result that in any case is small and not confirmed for 10th-graders. For mental

Table 1
Achievement scores obtained by 9th- and 10th-graders attending schools in North and South Italy * $p<.01$, ** $p<.001$.

	Group	9th grade								10th grade							
		N	M	SD	SEM	Pooled SD	Effect size	t test		N	M	SD	SEM	Pooled SD	Effect size	t test	
								df	t							df	t
Reading Comprehension	North	430	10.51	2.83	0.14	2.89	0.31	751	4.24**	310	13.18	3.65	0.21	3.62	0.43	621	5.41**
	South	323	9.61	2.96	0.16					313	11.61	3.58	0.20				
Arithmetic	North	445	6.79	2.83	0.13	2.76	0.27	675	3.28*	275	4.83	1.98	0.12	1.92	0.43	470	4.56**
	South	232	6.06	2.62	0.17					197	4.02	1.84	0.13				
Algebra	North									267	5.26	2.23	0.14	2.05	0.71	489	7.80**
	South									224	3.82	1.80	0.12				
Geometry and Measures	North	324	6.42	2.97	0.17	2.85	0.47	554	5.43**	242	4.14	1.71	0.11	1.55	0.44	423	4.51**
	South	232	5.09	2.66	0.17					183	3.46	1.31	0.10				
Arithmetic Problem-solving	North	235	5.19	2.50	0.16	2.44	0.25	358	2.25								
	South	125	4.58	2.32	0.21												
Reading Decoding Speed	North	152	5.41	0.99	0.08	1.01	0.36	459	3.68**	143	5.54	0.99	0.08	0.93	0.30	509	3.27*
	South	309	5.04	1.02	0.06					368	5.27	0.91	0.05				
Reading Decoding Errors	North	152	6.84	5.52	0.45	4.51	0.66	459	6.63**	143	7.24	5.14	0.43	4.09	0.84	509	9.03**
	South	309	3.88	3.92	0.22					368	3.80	3.61	0.19				
Mental Calculation Accuracy	North	82	4.50	1.77	0.19	1.91	0.35	355	2.76*	98	3.76	1.67	0.17	1.82	0.11	428	0.90
	South	275	3.84	1.95	0.12					332	3.56	1.87	0.10				
Mental Calculation Time	North	85	131.01	48.78	5.29	77.31	0.31	394	2.54	100	145.21	57.92	5.79	85.46	0.17	460	1.45
	South	311	106.97	83.38	4.73					362	130.53	91.58	4.81				
Arithmetical Facts	North	85	20.08	3.32	0.36	4.51	0.28	394	2.32	100	18.74	3.83	0.38	4.76	0.21	460	1.34
	South	311	18.80	4.78	0.27					362	17.74	4.98	0.26				

calculation, a trade-off is evident as the faster students were also less accurate—for example, in 9th-graders, the higher accuracy in mental calculation of the Northern children (0.35) should be seen against the higher speed in the Southern children (0.31).

As suggested by Rushton and coauthors (e.g., Rushton, Bons, Vernon, & Cvorovic, 2007; Rushton & Skuy, 2000; Rushton, Skuy, & Fridjhon, 2003), group differences in IQ can be better described on the basis of an analysis at item level. Despite the fact that it may be open to criticisms (for a debate see Rushton & Jensen, 2010; Wicherts & Johnson, 2009) the method can offer further information useful in the present context. With our present data set, it was possible to run analyses at this level only for the case of the 310 North Italy and 313 South Italy 10th-graders tested with the reading comprehension test (20 items). To test whether North–South differences are more pronounced on the more g-loaded items, we calculated the standardised differences between the percentages in the two groups of correct responses for each item, and item-total correlation (using point-biserial correlation, [r_{pb}]) for both groups. Moreover, we verified if item difficulties, measured by the percentage of correct responses between North and South for each item, were similar for the two groups (Pearson's $r=.93$; Spearman's $rho=.91$; $p<.001$). This suggests that the test measures the same construct in both groups. We then correlated standardised differences with both item-total correlations obtained with the Northern sample obtaining a relatively high correlation (Pearson's $r=.28$, ns; Spearman's $rho=.28$, ns), and with the Southern sample where the correlation was absolutely low (Pearson's $r=-.09$, ns; Spearman's $rho=.01$, ns). Altogether the observations collected with the MT-Advanced tasks, at least for North and South Italy samples, cast doubt on the generalizability of PISA data and on the appropriateness of using achievement scores to derive IQ differences.

1.2. INVALSI data

The latest measurements reported by INVALSI (2009a) give an overall description, which in many respects is different from that obtained from PISA (OECD, 2007). As the government-funded Italian national institute responsible for assessment of the learning outcomes of the Italian school system, INVALSI periodically produces reports presenting results of large-scale studies. Two recent reports (INVALSI, 2009a,b) provide information on such studies conducted with the same methodology as PISA, but focused on younger children (i.e., 2nd- and 5th-graders during school year 2008–2009—INVALSI, 2009a). Subjects were tested in language and mathematics in most cases by teachers of the same school, but not same class, under the supervision of external observers specifically trained for the project, who carried out direct control of the task administration to a subsample of randomly selected schools. The assessment involved 172,136 2nd-graders and 172,992 5th-graders, representing random subsamples within the schools selected by INVALSI to test populations representative of the population of the various regions of Italy. Table 2 presents the INVALSI study data, split into the three classically distinguished main areas, with Umbria, Tuscany, Marche, Lazio as Regions of Central Italy, and the other Regions in the North and South, respectively.

Table 2 shows the mean percentages (and standard deviations) of correct responses (and confidence intervals, *CI*, for $p=.95$) obtained by children of North, Central and Southern Italy. As seen, differences are present, but less dramatic than in the PISA study. Northern Italy children are superior in language, with Central Italy children slightly below the lower Confidence Interval of Northern children in 2nd-grade (but not in 5th-grade) and with Southern Italy children below both the other groups in both cases. Instead, the inferiority of Southern children is less evident for maths

Table 2
Mean percentages of correct response, and Confidence Intervals, for the INVALSI study, given by Italian 2nd- and 5th-graders, of North, Central and South Italy, in Italian and maths.

	Group	2nd grade					5th grade				
		N	M	SD	LL	UL	N	M	SD	LL	UL
Italian	North	20,253	67.3	21.3	66.7	68.0	20,154	63.1	13.2	62.6	63.5
	Central	9278	66.3	21.4	65.2	67.3	9229	62.9	13.6	62.2	63.6
	South	13,808	61.6	22.5	60.5	62.7	14,133	61.1	15.6	60.1	62.1
Math	North	20,253	55.2	17.6	54.6	55.8	20,154	58.9	15.4	58.4	59.4
	Central	9279	54.3	17.7	53.5	55.0	9229	57.4	15.8	56.4	58.4
	South	13,823	54.8	19.7	53.4	56.1	14,133	55.1	17.0	53.9	56.3

Note: Italian Regions. Northern: Piedmont, Aosta Valley, Liguria, Lombardy, Trentino-Alto Adige, Veneto, Friuli Venezia Giulia, Emilia-Romagna; Central: Tuscany, Umbria, Marche, Lazio; Southern: Abruzzo, Molise, Campania, Apulia, Basilicata, Calabria, Sicily, Sardinia. Confidence Intervals ($p=.95$). LL = Lower Limit; UL = Upper Limit.

and concerns only 5th-graders. We computed the effect sizes (Cohen's d) for the differences between Northern and Southern Italian children which were, for Language, respectively .26 for second-graders and .14 for fifth-graders and, for Mathematics, respectively .02 and .24.

Two other interesting observations that can be made regarding the INVALSI (2009a) data concern the larger variability (in particular between-schools) and the greater polarisation found in Southern Italy. Variance decomposition was obtained from the ANOVAs separated for Grade (2nd and 5th) and Subject (Italian, Math) with the following design: Region (North, Centre, South) × Schools: this allowed calculation of both within- and between-schools variance and the corresponding percentages with respect to total variance.

In our view, the between-classes variability reflects the strong effects due to the different schools and different sociocultural contexts present in South Italy. For example, in 2nd grade, variance in language between schools is very low in Northern Italy (less than 5% of overall variance), lower than 9% in Central Italy, but 17.5% in the South. In other words, Southern Italy schools, despite having the same curricula and same ethnicity, present large differences, which are even larger considering just maths. The South generally presents a greater variance, with a sizable part due to differences between schools, 28.3% vs. values lower than 10% in Northern and Central Italy. Between-classes variance is even higher for 5th-graders (25.2% for language, 37.1% for maths).

Polarisation refers to the fact that Southern children have a higher probability of occupying both the lowest parts and the highest parts of the achievement distribution. For example, the percentages of Southern children below 10% percentile and above 90% percentile are some 2% higher than for the remainder of Italy, showing that in South Italy it is possible to find not only more weaknesses but also more talented children.

As the INVALSI data set also held information on 8th-graders (testing 1307 classes—an estimated 26,150 children) and for various regions, we also examined the data for 8th-graders and distinguished achievement by Region. These data should be considered with caution, as the administered tasks had some problems, and different Regions were represented to different extents in view of their sizes, and were also affected by special contextual situations (e.g., in Bolzano and Aosta Valley some children present incomplete bilingualism). Table 3 shows these data together with the most recent available (Eurostat NewsRelease, 2010) on Italian regional annual income in Purchasing Power Standard (PPS; see Le Gallo, 2004). The PPS is an artificial currency that takes into account differences in local price levels. The unit allows meaningful volume comparisons of economic indicators over countries. Aggregates expressed in PPS are derived by dividing aggregates in current prices and national currency by the respective Purchasing Power Parity. PPS thus takes better account than other income measures of the actual financial possibilities of people living in different geographical areas. Note that, in any case, PPS offers information that very largely overlaps with the mean income measure used by the previous study (Lynn, 2010): for example, for the Italian regions, the two measures have a very high Pearson's correlation ($r=.99$).

A brief glance at Table 3 shows that the correspondence between income and achievement is not as clear-cut as that suggested by Lynn (2010). For example, Lombardy—the Region of Milan and the wealthiest in Italy—shows relatively poor reading and maths performance in 2nd and 5th grades.

This modest (and unclear) relationship between income and achievement in primary school is also confirmed by the correlations in Table 4. Instead, for 8th-graders the relationship is strong, though weaker than that reported by Lynn (2010) for 15-year-olds.

1.3. Raven's data

Lynn (2010) mentions earlier data obtained by Pruneti (1985) on Raven's Coloured Progressive Matrices (CPM; Raven, Court, & Raven, 1992), without considering the more robust and recent normative data reported in the standardisation of Belacchi et al. (2008). Although this latter standardisation lacked a differential examination of the performance of children from different areas of Italy, it offers a preliminary result of interest (i.e., typical mean performance of Italian children).

For this present analysis, we have focused on a particular age, intermediate between 2nd and 5th grades, where sufficient numbers of Northern and Southern children could be compared. Before comparing North and South Italy, we examined the general national trend emerging from the new standardisation. Table 5 presents a comparison between the recent Italian norms (for children in groups 8.6–9 and 9–9.5, combined) and those for other countries. These findings bring to light a number of ambiguities, since children of different countries may present different outcomes that seemingly

Table 3
Regional PPS and mean percentages of correct scores on Italian and maths tests in 2nd, 5th and raw scores for 8th grades (only regularly attending students) from INVALSI.

	PPS	2nd Grade				5th Grade				8th Grade			
		Italian		Math		Italian		Math		Italian		Math	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Piedmont	28,300	67.0	21.4	55.3	17.8	63.6	13.2	59.3	15.5	27.8	6.2	18.0	5.2
Aosta Valley	29,500	71.5	20.3	55.6	16.6	63.0	12.8	57.0	14.6	29.6	5.7	17.8	4.9
Lombardy	33,600	68.9	20.8	55.4	17.4	63.2	12.9	58.4	15.0	28.1	7.0	18.8	4.7
Liguria	26,600	66.1	21.7	53.7	17.6	62.6	13.7	57.0	15.7	27.6	6.7	18.2	4.9
Trento	30,400	66.1	21.5	54.3	18.1	61.8	13.4	59.3	14.8	27.4	6.4	18.6	4.4
Bolzano (Italian)	33,500	61.7	24.1	50.5	18.9	56.1	12.1	51.7	15.2	26.8	6.8	17.7	5.1
Veneto	30,300	65.7	21.5	55.5	17.6	63.1	13.1	60.4	15.6	27.8	6.6	18.5	4.7
Friuli Venezia Giulia	29,000	68.2	21.5	55.1	17.8	63.2	13.2	59.9	16.1	28.3	7.1	19.0	4.5
Emilia-Romagna	31,900	66.6	21.9	55.2	17.8	62.8	13.8	58.5	15.5	27.5	6.6	18.1	5.0
Tuscany	28,100	66.3	21.5	54.5	17.3	64.4	13.7	59.8	16.6	28.0	7.0	18.8	4.8
Umbria	24,100	67.4	21.0	56.2	17.9	63.9	13.5	59.8	15.6	28.1	7.7	16.8	6.8
Marche	26,300	68.7	20.2	56.5	17.8	65.3	13.6	60.7	16.0	28.5	6.3	19.0	4.5
Lazio	30,500	65.5	21.6	53.3	17.7	61.3	13.3	54.7	14.9	27.7	6.7	17.9	4.9
Abruzzo	21,200	64.9	21.5	54.7	18.5	62.4	13.7	55.3	15.0	26.1	7.7	17.4	5.6
Molise	19,400	66.9	21.1	56.2	17.6	64.5	13.9	57.2	16.0	27.8	5.9	17.3	5.3
Campania	16,400	60.5	23.2	56.0	20.8	62.4	15.8	57.2	18.0	24.7	9.5	14.4	7.1
Apulia	16,600	63.4	22.0	56.7	19.2	62.3	15.5	57.3	16.6	25.5	8.8	16.9	6.1
Basilicata	18,700	65.2	20.7	56.5	18.1	62.9	14.1	56.9	15.2	26.4	7.6	16.7	6.7
Calabria	16,400	63.7	22.4	57.4	21.2	63.0	15.9	57.4	17.4	24.1	9.9	15.1	6.9
Sicily	16,400	59.3	22.0	50.8	18.3	57.9	15.6	50.4	15.9	24.8	8.3	15.1	6.3
Sardinia	19,500	62.5	22.6	53.2	18.1	59.0	14.6	51.6	15.4	26.4	6.3	16.3	5.5
Italy		65.0	21.9	54.9	18.4	62.3	14.3	57.1	16.2	26.8	7.7	17.2	5.8

cannot be explained considering the Flynn (1987) effect alone. On the basis of these data, Italian children do not appear particularly outstanding or advantaged through being a recent cohort. In fact, comparison of CPM Italian norms with four other banks of data gathered from 1954 to 2008 revealed no substantial changes (Belacchi et al., 2008, p. 54) although it should be noted that the Italian children were slightly younger than the other groups. However, there are also marked differences between the various countries, for example the French children being particularly bright, while British and Italian children show lower performance levels. These results suggest that comparisons between different test administrations should be always viewed with caution.

Using the data collected for producing the recent Italian CPM normative data allowed comparison of the administration for the different Regions of Italy. We were in fact able to establish the location of the 747 children aged between 7.6 and 9.11 in the normative sample and living in either Northern Italy (Lombardy, Emilia-Romagna, Friuli Venezia Giulia, and Veneto) or Southern Italy (Abruzzo and Apulia). Table 6 shows the mean raw scores obtained by the different groups. A 5 × 2 ANOVA (Age [7.6 to 7.11, 8 to 8.5, 8.6 to 8.11, 9.0 to 9.5, and 9.6 to 9.11] × Geographical Area [North, South]) showed a significant effect of Age, $F(4, 737) = 14.00$, $p < .001$, $\eta_p^2 = .071$, but no significant effect of Geographical Area, $F(1, 737) = 1.78$, $p = .18$, $\eta_p^2 = .002$, and no significant interaction between age and geographical area, $F(4, 737) = .16$, $p > .05$, $\eta_p^2 = .001$. Note that the Italian norms associate raw scores with percentile scores but not with IQs. Both groups obtained mean scores very close to 50th percentile. Based on these norms, the mean scores obtained by Northern Italian children correspond approximately to percentiles: 53, 51, 53, 54, 57 and those of Southern Italian children to percentiles: 54, 45, 50, 48, and 50 which can be roughly estimated as respectively mean IQs of 101, 101, 101, 102 and 103 for Northern Italian children and of 102, 98, 100, 99, 100 for Southern Italian children.

Table 4
Pearson's correlations (Spearman rho in parentheses) between regions PPS and mean scores obtained in Italian and maths in the INVALSI achievement tests (see data in Table 3).

Grade	Italian	Maths
2nd	.56** (.49*)	−.25 (−.41)
5th	.03 (.03)	.30 (.24)
8th	.74** (56**)	.82** (.72**)

Note: * $p < .05$, ** $p < .01$.

2. Discussion

Our analysis of the data described confirms our doubts as to the generality of the PISA data and the validity of using them for making generalisations on intelligence and its relationship with other variables, as Lynn (2010) has done. The MT-Advanced tasks and Raven Coloured Matrices were carried out on relatively small samples (and the psychometric properties of some tasks are not particularly robust), and consequently the data obtained should therefore be viewed with caution, but they were nevertheless consistent both internally and also with the large INVALSI study data. Together, these data support our perplexities, set out below under seven points.

(1) As already suggested by various authors (Duru-Bellat & Suchaut, 2005), the PISA data, collected to provide estimates of the effective school performance by children from different countries, is little suited for

Table 5
Percentiles observed for the standardisation of Raven's Coloured Progressive Matrices in various countries. Ages in parentheses.

PR	Italy	GB	France	Switzerland	Spain	E. Germany	W. Germany
	(8.6–9.5)	(8.9–9.2)	(8.9–9.2)	(8.9–9.2)	(8.9–9.2)	(8.9–9.2)	(8.9–9.2)
90	32	31	34	34	33	32	33
75	30	28	33	32	29	30	31
50	26	25	30	28	26	27	27
25	23	20	27	25	22	23	22
10	18	17	24	23	17	20	18

Note: Italy, data were extrapolated from Belacchi et al. (2008); GB, from Raven, Raven, and Court (1998 Table 10, p. 56); France, from Raven et al. (1998, Table 10, p. 56); Switzerland from Raven et al. (1998, Table 15, p. 58); Spain, from Raven et al. (1998, Table 16, p. 59); East Germany, from Raven et al. (1998, Table 17, p. 59); West Germany from Raven et al. (1998, Table 13, p. 57). PR = Percentile Rank.

providing comparative estimates on effective intellectual ability, since they also reflect different contexts that are difficult to compare. In fact, if these contextual factors are taken into account, the differences between North and South Italy tend to disappear (Bratti et al., 2007). The fact that contextual factors are critical in the Italian PISA data is also supported by the greater between-schools variance and the corresponding percentages with respect to the total variance we found in the Southern sample, corresponding to the well-known greater socio-economic heterogeneity of South Italy with respect to the North (Felice, 2007). One aspect in particular, already considered by Lynn, concerns level of family education (schooling). If we consider educational level at 1971—the index closest to the educational level of the parents of those tested under the PISA study—we in fact see that this correlates very strongly (.87) with the children's estimated IQ. It is certainly true that both indices could be reflecting just a deeper-rooted connection between genetic factors, intelligence and educational level (Bartels, van Beijsterveldt, & Boomsma, 2009; Lynn, 2010) but evidence (Cliffordson & Gustafsson, 2008) suggests that educational level may be the cause of intellectual capacity rather than the effect. Wicherts, Dolan, and van der Maas (2010) have raised similar arguments concerning the postulated differences in IQ between Europeans and sub-Saharan Africans.

(2) Administration of group tests in reading and maths, similar to the PISA tests but excluding students who had clearly not collaborated, revealed less marked differences, typically of the order of .4 SDs of those found in the PISA study (typically around 1 SD). Furthermore, a tentative analysis of the Reading Comprehension items suggested that the North–South differences were not related with those items more critically measuring the construct. Many factors could have produced the differences between MT and PISA observations, including the different properties of the tests and the different procedures for selecting subjects, but considering only students who collaborated with the procedure could be a critical factor. Although low motivation could also result from low intelligence, the fact that many students in the PISA study did not take the task seriously could be due to other factors than intelligence, such as attributional beliefs (Moè, & Pazzaglia, in press), as well as adding noise to the data.

(3) The simultaneous administration of individual reading and maths tests, which involved—with only a few exceptions—the full complement of students from the classes involved, suggested that the group differences could be partly due to biases present in the group administration. The North–South differences largely disappear and speed–accuracy trade-off phenomena arise; their generality as regards the North–South comparison needs closer examination. In particular, children from Southern Italy seem to be slower but more accurate in reading, and faster but more prone to mistakes in maths.

(4) Analysis of the performance of children younger than those involved in the PISA study further confirms our conclusion that the differences in achievement between North and South have been overemphasised, and are based on conclusions obtained regarding 15-year-olds, which cannot be generalised to younger children. These differences are fairly modest and appear in only certain cases, suggesting that the largest differences found in the group testing of children who are older (i.e., with more schooling) could also be due to the school effect itself. There is much data to suggest that school quality in the South is actually inferior and that this quality relates to level of learning. In addition, the measurements from the various studies may be influenced by the larger between-schools variability

Table 6
Mean raw scores obtained on the Raven's Coloured Progressive Matrices (and standard deviations, SD) by children from North and South Italy, aged 7.6 to 9.11.

Age		North	South	Effect size
7.6–7.11	N	48	61	−.03
	M	22.98	23.13	
	SD	6.21	5.03	
8.0–8.5	N	69	50	.13
	M	24.17	23.48	
	SD	5.43	4.81	
8.6–8.11	N	110	81	.11
	M	25.68	25.04	
	SD	5.39	5.90	
9.0–9.5	N	57	109	.15
	M	27.14	26.35	
	SD	4.60	5.54	
9.6–9.11	N	54	108	.18
	M	27.37	26.51	
	SD	4.71	4.64	

present in the South. As noted above, the between-schools and between-classes variance has positive relationship with overall average of test scores—this may therefore also partly explain the lower results of the South compared with the rest of Italy.

Even accepting the hypothesis that IQ arises partly through school and environmental stimulation, it is nevertheless true that a genetic base of IQ has been shown (Plomin & Petrill, 1997). In particular, literature addressing ethnic and geographical differences in IQ (e.g. Lynn, 2010) seems to be concerned with the part of IQ difference not due to environment. It is possible that some genetic differences emerge more clearly with age: in fact increases in ethnic differences in mean IQ with age have been reported and genetically interpreted (Jensen, 1998). However, under the environmental view—focused on the effects of schooling—these latter effects should be more evident in older students. It therefore seems unwise to derive an IQ measure and compare it across different contexts from estimates of school learning, in particular those obtained from older children who have experienced different levels of environmental stimulation: a more judicious approach would be to consider younger children.

(5) Comparing the levels of learning at different grades, the relationship between learning and regional differences seems particularly low at early grades, being most valid for older children. The relationship between the PISA results and national income is well established; the most recent figures show that per capita gross domestic product (GDP) explains 28% of the variability in science results between countries (OECD, 2007, p. 59). In the present study we have favoured a more direct measure of purchasing power (PPS), but a subsequent correlation showed that differences in the results cannot be attributed to our use of an income measure different from that adopted by Lynn (2010), as the two measures highly correlated. Jensen (1969) argued that quality of instruction is affected by the underlying level of intelligence. However the fact that the income/school–success relationship is larger for older children could also be due to the fact that this relationship is not primitive, but a consequence of the lower quality of educational institutions in the poorer Regions.

(6) The idea that more excellence is found in the North is not confirmed by our data. Measurement of individuals who are outstandingly bright in the various fields of arts and sciences is always difficult, but seems to be particularly sensitive if used to give an estimate of the intelligence of various populations; instead, it might be useful to look directly at how frequently outstanding children appear. The Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione (INVALSI) (2009a) data set, involving a very large number of children, allows preliminary observations here, and shows that the percentage of children obtaining particularly high scores is higher in the South, rather than lower.

(7) If an estimate of IQ is sought from a classic intelligence test, less sensitive to cultural and school conditioning than learning tests (Raven et al., 1992), then the North–South differential between children appears to be still smaller. At the same time, there is further confirmation that international comparison—if implausible conclusions are to be avoided (for example that French infants have a particularly high IQ) is affected by how the tests are administered, leading to findings which must be cautiously used when drawing conclusions about intelligence differences between populations.

We are aware that some of our conclusions are based on data that might not appear as fully robust as the PISA data, and that different implications could be advanced for the same data. However, we believe the data we present and the conclusions derived may be useful in reappraising both the generalisability of the PISA data, and also their possible use for generalisations that do not take account of contextual variables. Furthermore, our evidence confirms the warning that national IQ data should be used very cautiously as support to biological theories of intelligence (Wicherts, Borsboom, & Dolan, 2010). We recognise that our study should be consolidated with a broader dataset of greater reliability. Nevertheless, a main conclusion remains that generalisations bringing into play very important assessments of entire populations should be approached with care and awareness.

References

A'Hearn, B. (1998). Institutions, externalities, and economic growth in southern Italy: Evidence from the cotton textile industry, 1861–1914. *The economic history review*, *51*, 734−762.

Bartels, M., van Beijsterveldt, C. E. M., & Boomsma, D. I. (2009). Breastfeeding, maternal education and cognitive function: A prospective study in twins. *Behavior Genetics*, *39*, 616−622.

Belacchi, C., Scalisi, T. G., Cannoni, E., & Cornoldi, C. (2008). *CPM—Coloured progressive matrices standardizzazione Italiana.* Florence, Italy: Giunti O.S.

Bratti, M., Checchi, D., & Filippin, A. (2007). Territorial differences in Italian students' mathematical competences: Evidence from PISA 2003. *IZA Discussion Paper* (N. 2603).

Cliffordson, C., & Gustafsson, J. E. (2008). Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continuous variation in age and length of schooling. *Intelligence*, *36*, 143−152. doi:10.1016/j.intell.2007.03.006

Cornoldi, C., Friso, G., & Pra Baldi, A. (2010). *Prove MT avanzate di lettura e matematica 2 per il biennio della scuola secondaria di II grado. [Advanced MT 2 tests—Advanced MT tests of reading and math for 9th and 10th grades].* Florence, Italy: Organizzazioni Speciali.

Cornoldi, C., Pra Baldi, A., & Rizzo, M. (1991). *Prove avanzate MT di comprensione nella lettura [Advanced MT tests of reading comprehension].* Florence, Italy: Organizzazioni Speciali.

Duru-Bellat, M., & Suchaut, B. (2005). Organisation and context, efficiency and equity of education systems: What PISA tells us? *European Educational Research Journal*, *4*, 181−194.

Eurostat NewsRelease (18, February, 2010). Regional GDP per inhabitant in 2007. Retrieved from http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/1-18022010-AP/EN/1-18022010-AP-EN.PDF

Felice, E. (2007). *Divari regionali e intervento pubblico: per una rilettura dello sviluppo in Italia [Regional differences and public intervention: rereading Italian development].* Bologna, Italy: Il Mulino.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*, 171−191. doi:10.1037/0033-2909.101.2.171

Herrstein, R. J., & Murray, C. A. (1994). *Bell curve.* New York, NY: Free Press.

Howe, M. (1997). *IQ in question: The truth about intelligence.* London, UK: Sage.

Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione (INVALSI) (2009a). *Rilevazione degli apprendimenti scuola primaria [Assessment of primary school achievement].* Rome, Italy: Author.

Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione (INVALSI) (2009b). *Esame di Stato. Prova Nazionale [National examination].* Rome, Italy: Author.

Jencks, S. (1972). *Inequality.* London, UK: Penguin.

Jensen, A. R. (1969). How much can we boost I.Q. and scholastic achievement? *Harvard Educational Review*, *39*, 1—23.

Jensen, A. R. (1985). The nature of the Black and White difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, *8*, 193—263.

Jensen, A. R. (1998). *The g factor.* Westport, CT: Praeger.

Jimenez, J. E., & Garcia de la Cadena, C. (2007). Learning disabilities in Guatemala and Spain: A cross-national study of the prevalence and cognitive processes associated with reading and spelling disabilities. *Learning Disabilities Research & Practice*, *22*, 161—169.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology*, *85*, 103—119.

Le Gallo, J. (2004). Space-time analysis of GDP disparities among European Regions: A Markov chains approach. *International Regional Science Review*, *27*, 138—163. doi:10.1177/0160017603262402

Lennon, R. T. (1950). The relation between intelligence and achievement test results for a group of communities. *Journal of Educational Psychology*, *41*, 301—308.

Lindgren, D., De Renzi, E., & Richman, L. C. (1985). Cross-national comparisons of developmental dyslexia in Italy and the United States. *Child Development*, *56*, 1404—1417.

Lihshing, L. W. (2010). Disattenuation of correlation due to fallible measurement. *Newborn and Infant Nursing Reviews*, *10*, 60—65.

Lynn, R. (2010). In Italy, north–south differences in IQ predict differences in income, education, infant mortality, stature, and literacy. *Intelligence*, *38*, 93—100.

Moè, A., & Pazzaglia, F. (in press). Beyond Genetics in Mental Rotation test performance. The power of effort attribution. *Learning and Individual Differences*. doi:10.1016/j.lindif.2010.03.004

Organisation for Economic Co-operation & Development (OECD) (2003). *Education at a glance.* Paris, France: Author.

Organisation for Economic Co-operation & Development (OECD) (2007). *PISA 2006 competencies for tomorrow's world.* Paris, France: Author.

Organisation for Economic Co-operation & Development (OECD) (2009). *PISA 2006 technical report.* Paris, France: Author.

Plomin, R., & Petrill, S. A. (1997). Genetics and intelligence: What's new? *Intelligence*, *24*, 53—77.

Pruneti, C. (1985). Dati normativi del test PM 47 Coloured su un campione di bambini italiani. (Normative data in an Italian sample for Coloured PM 47). *Bollettino di Psicologia Applicata*, *176*, 27—35.

Quintano, C. (2007, November). Dati di qualità per una scuola di qualità (Quality in data and in school). SisMagazine. Retrieved from http://www.sis-statistica.it/magazine/spip.php?article48

Raven, J. C., Court, J. H., & Raven, J. (1992). *Raven's coloured progressive matrices.* Oxford, UK: Oxford Psychologist Press.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual, Section 2 (Coloured Progressive Matrices).* Oxford, UK: Oxford Psychologist Press.

Rushton, J. P., Bons, T. A., Vernon, P. A., & Cvorovic, J. (2007). Genetic and environmental contributions to population group differences on the Raven's Progressive Matrices estimated from twins reared together and apart. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *274*, 1773—1777.

Rushton, J. P., & Jensen, A. R. (2010). Race and IQ: A theory-based review of the research in Richard Nisbett's Intelligence and How to Get It. *The Open Psychology Journal*, *3*, 9—35.

Rushton, J. P., & Skuy, M. (2000). Performance on Raven's Matrices by African and White university students in South Africa. *Intelligence*, *28*, 251—265.

Rushton, J. P., Skuy, M., & Fridjhon, P. (2003). Performance on Raven's Advanced Progressive Matrices by African, East Indian, and White engineering students in South Africa. *Intelligence*, *31*, 123—137.

Santello, F. (January, 2009). *Identification and estimation of the effects of school quality on the achievement of Italian students*. Paper presented at the Meeting of Italian Congress of Econometrics and Empirical Economics, Ancona, Italy. Retrieved from http://www.cide.info/conf/papers_2009.php

Stanovich, K. E. (1991). Discrepancy definitions of reading disability: Has intelligence led us astray? *Reading Research Quarterly*, *26*, 7—29.

Wicherts, J. M., Borsboom, D., & Dolan, C. V. (2010). Why national IQs do not support evolutionary theories of intelligence. *Personality and Individual Differences*, *48*, 91—96.

Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. J. (2010). A systematic literature review of the average IQ of sub-Saharan Africans. *Intelligence*, *38*, 1—20.

Wicherts, J. M., & Johnson, W. (2009). Group differences in the heritability of item and test scores. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *276*, 2675—2683.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3—29.

Zorman, M., Lequette, C., & Pouget, G. (2004). Dyslexies: Intérêt d'un dépistage et d'une prise en charge précoce à l'école. In M. N. Metz-Lutz, E. Demont, C. Seegmuller, M. de Agostini, & N. Bruneau (Eds.), *Développement cognitif et troubles des apprentissages: Evaluer, comprendre, rééduquer et prendre en charge* (pp. 245—270). Marseille, France: Solal.