Differences in achievement not in intelligence in the north and south of Italy: Comments on Lynn (2010a, 2010b)

Antonella D’Amico a,⁎, Maurizio Cardaci a, Santo Di Nuovo b, Jack A. Naglieri c

a Dipartimento di Psicologia, Università degli Studi di Palermo, Italy
b Dipartimento dei Processi Formativi, Università degli Studi di Catania, Italy
c University of Virginia, United States

1. Introduction

In a recent paper published in the journal Intelligence Lynn (2010a) argued that north–south differences in Italians’ IQ scores predict differences in income, education, infant mortality, stature and literacy. Lynn’s also wrote that this IQ difference “has a genetic basis going back many centuries, and hence predicts the social and economic differences documented in the nineteenth century up to the present day” (pp. 99). His paper evoked a strong reaction from the Italian scientific community both through internet (see: http://www.aipass.org/node/319) and in the same journal (Beraldo, 2010; Cornoldi, Belacchi, Giofre, Martini, & Tressoldi, 2010; Felice & Giugliano, 2010). Lynn (2010b) replied with new arguments that, again, seem quite questionable. In the present paper we discuss several theoretical and methodological aspects which we regard as flaws of Lynn’s studies. Moreover, we report scores of southern Italian children on Raven’s Progressive Matrices and a north–south comparison for the PASS theory of intelligence as measured by the Cognitive Assessment System (Taddei & Naglieri, 2006). Both results reveal similar levels of performance of northern and southern Italian children in fluid intelligence and PASS (Planning, Attention, Simultaneous, and Successive) cognitive abilities.

1.1. The measurement of IQ

The question of the nature and measurement of intelligence has been a topic of considerable interest in Psychology in the last century, and it is not our aim to review the literature about this issue (a good review of the field is provided by Deary, Penke, & Johnson, 2010; Hunt, 2011). It is important, however, to revisit a few aspects that should always be considered in studying intelligence but especially in regard to Lynn’s selection of data upon which he has made his statements.

It is difficult to measure intelligence without considering the influence of social and cultural variables. Indeed, scores on verbal and quantitative test questions, on instruments such as the Wechsler (2003) or Stanford–Binet (Roid, 2003) scales, are strongly influenced by linguistic skills and related to educational quality. For these reasons, measures that exclude language were developed, such as the Raven’s Progressive Matrices (Raven, 1954) or the Cattell (1949), for assessing cognitive ability in a way that is minimally influenced by literacy, education and informal learning. Although the role of environmental conditions may be never totally controlled, individual, regional or national differences in IQs should be made with consideration of these factors. Moreover, great caution is needed when considering the issue of collective genetic differences in intelligence (e.g., Wicherts, Borsboom, & Dolan, 2010).

1.2. Differences in achievement not in intelligence

Lynn’s (2010a) estimate of IQ was based on the 2006 British PISA (Program for International Student Assessment), an internationally standardized assessment administered to 15 year olds in schools, that found higher scores for students in northern Italy when compared to students in the south. PISA tests, however, were developed to measure achievement and not intelligence. In fact, the aim of PISA is to measure “how far students near the end of compulsory
education have acquired some of the knowledge and skills that are essential for full participation in society” (for more information, see the site www.pisa.oecd.org).

Differences in scholastic achievement of Italians have been documented by Cornoldi et al. (2010) as well as by the Italian INVALSI (the National Institute for the Assessment of Educational and Vocational System). Cornoldi et al. (2010) used the MT-Advanced tasks (Cornoldi, Pra Baldi, & Rizzo, 1991) and demonstrated that more accurate methodological controls reduced these differences. INVALSI’s (2009) results showed that, in fifth grade, pupils in the north achieve better than children in the south, but there are no statistical differences between achievement of north and south pupils in the second grade. Moreover, a further INVALSI study by Campodifiori, Figura, Papini, and Ricci (2010) found high variability between performances of children belonging to different schools of the same southern towns. Both these results can be explained by the impact of socio-economic factors on scholastic achievement.

Nevertheless, Lynn (2010a) uses achievement tests as “proxies for Intelligence” (pp. 95) adopting the logic that educational attainment and intelligence are highly correlated (from r = 0.5 to r = 1.0) across nations (Lynn & Meisenberg, 2010; Lynn & Mikk, 2007). However, in his studies it is not clear what kind of IQ tests has been used, and the other factors affecting achievement such as school quality, socio-cultural level, and so on, are not controlled.

1.3. Correlation relationships discussed as causality relationships

It is widely known and accepted that a correlation coefficient describes the degree of relationship between two variables. However, two variables may correlate highly, but they may be different from each other. It is also possible that changes in the variables being studied are influenced by some other unobserved variable. Finally, correlation does not assume causality.

Against such universally shared methodological rules, Lynn (2010a) discusses association among variables as if they are equivalent or in a simple unilinear causal relationship.

1.4. Regions as “subjects”

Lynn (2010a) stated that:

“data have been assembled for 12 Italian regions for mean IQ, average per capita income in Euros for 1970 and 2003 (…), percentages of the populations that were literate in 1880 (…) statures of military conscripts born in 1855, 1910, 1927 and 1980 (…) infant mortality 1955–57 and 1999–2002 (…), years of education in 1951, 1971 and 2001 (…) and latitude (…). The regional IQs have been calculated from the 2006 PISA (Program for International Student Assessment) study of reading comprehension, mathematical ability, and science understanding administered to 15 year olds in 52 countries (OECD, 2007)” (pp. 95).

Thus, Lynn (2010a) uses regions as “subjects”, therefore scores of “subject-region” correspond to the average measure of all the subjects that have been tested in that region. Consequently, participants of the study are always different individuals, of different age cohorts, sharing only the common aspect of living in the same Italian region. For instance, scholastic achievement (labeled by Lynn as IQ), is collected in 2006 in 15 year olds, while stature of subjects is collected in 1885, in an unknown number of subjects of unknown age. The same could be said for the other variables. It also means that the correlational study by Lynn (2010a) is performed on 12 “subjects”, and this is not statistically rigorous, as already stated also by Beraldo (2010).

Another problem with Lynn’s study refers to the representativeness of the sample used, since PISA results were only based on 15 year olds attending school. These subjects are not representative of the Italian population, because achievement levels change during the academic career. Moreover, data are collected only on the part of youth that attends school, while not all young persons attend school and not all young persons attend school regularly (Rindermann, 2007).

Finally, Lynn affirms that the regional differences in IQ (actually, differences in scholastic achievement) strongly reflect genetic differences between Italian population of north and south Italy. However, students who attend schools in the north of Italy, are not necessarily born in the north of Italy, from northern parents, and do not necessarily have “northern genes”.

On the basis of the points discussed so far, a significant adjustment should be made to the title of Lynn’s (2010a) paper, that should read: “In Italy, differences in scholastic achievement among 15 years old attending schools in the regions of north and south are associated with differences in income, education, infant mortality, stature, and literacy, measured in different populations that lived in the same regions in the period between the 1880 and 2001”. This title is really difficult to understand but it is accurate in describing what Lynn has found in his study.

1.5. Measuring intelligence using unvalidated tests

In his more recent paper, Lynn (2010b) reports further evidence of the lower IQs of southern Italians. The first is the report of an intelligence test given to a sample of 50,000 individuals who self-administered the test over the internet on www.sitozero.it. This is a commercial site with an inadequate description of the psychological tests used, with a considerable amount of advertisements and without any control of scientific and methodological issues. We do not consider these non-scientific data to be suitable for making assumptions about IQs.

1.6. Intelligence scores and Flynn effect

Lynn (2010b) uses data from several studies on Raven’s test (Pruneti, 1985; Pruneti, Fenu, Freschi, & Rota, 1996; Tesi & Young, 1962) and Cattell Culture Fair test (Buj, 1981; Pace & Sprini, 1998). None of the studies used the same age groups and none were aimed at comparing IQs across regions of Italy.

Moreover, Lynn (2010b) did not consider the calculation of IQs made by the authors, but rather he recalculated the IQs scores in light of the well known and controversial (Colom, Lluis-Font & Andrés-Pueyo, 2005) Flynn effect (2007), described as a general increase of intelligence scores over the worlds in the last 50 years. So, for instance, an IQ of 99 collected in 1960, was increased by 4 points considering the Flynn effect = 4 of the Italian IQ in the years 1960–79.

Such procedure is questionable, as also Hagan, Drogin, and Guilmette (2008) pointed out. Indeed, different studies demonstrated that the Flynn effect is concentrated in the lower half of the normal distribution or in undeveloped countries (Colom et al., 2005), whereas a possible stagnation of IQ scores in developed ones is currently under debate (Teasdale & Owen, 2005; 2008).

The best way to study regional differences is to compare subjects from the same age cohort who live in different geographical regions using the same test. This was conducted by Cornoldi et al. (2010). The authors drew from a larger standardization sample of Raven Coloured Progressive matrices made by Belacchi, Scalisi, Cannoni, and Cornoldi (2008) involving a group of 747 children belonging to 5 age groups living in northern or southern Italy. Then, they compared their CMP scores through a 5 × 2 ANOVA age × geographical area. Results showed a significant effect of age, but no significant effect of geographical area. Lynn (2010b) criticized these results arguing that the Belacchi et al.’s Italian standardization of CMP “is clearly defective” (pp. 454) because the authors failed to detect the expected
Flynn effect. Lynn averaged the mean raw scores of three out of five age groups described by Cornoldi et al. (2010) and compared those obtained scores to the British 1982 standardization. Although he found that the (British) scores of northern children was 3.7 IQ points higher than the score of Southern children we believe that the use of the British 1982 standardization is questionable, and it is unclear why Lynn excluded two age groups from comparisons.

2. New evidence against the north south differences in IQs

With the aim to contribute to the study of regional differences in IQs, we obtained two new sources of evidence based on the direct assessment of IQs in children of different Italian regions, using measures of intelligence that do not contain highly academic content.

3. Study 1: Raven's data

3.1. Method

Raven's CPM data about a group of Sicilian children collected for different research purposes (D'Amico & La Porta, 2010; D'Amico, Lipari, & La Porta, 2008; D'Amico & Passolunghi, 2009), were compared to CPM's data from Belacchi et al. (2008) and Cornoldi et al. (2010).

3.2. Participants

Participants of the study were 288 children (146 males, 142 females, aged between 8.5 and 9.6), attending the fourth and fifth grade of primary schools situated in suburban areas of Palermo, Sicily. None of the children was receiving special education services or had documented brain injury, or behavioral problems. As reported by teachers, none of the children were from groups with cultural disadvantage. The data of children that attended Sicilian schools were collected between the years 2007 and 2008, so they belonged to the same age cohort described by Belacchi et al. (2008) and Cornoldi et al. (2010).

3.3. Materials and procedure

Children completed the CPM test at school, in group sessions under the supervision of trained administrators. In order to compare our CPM scores with Cornoldi et al. (2010) ones, the children were grouped in the same age groups used by authors (age 8.6–8.11, 11 males, 15 females; age 9.0–9.5, 78 males, 75 females; age 9.6–9.11, 57 males, 52 females), and the raw scores of each age group were computed. Then, mean Raven's raw scores of children from Sicily were compared to mean raw score of children from northern and central-south of Italy, recently published by Cornoldi et al. (2010).

3.4. Results and discussion

Descriptive statistic of each age group is reported in Table 1. There is a slight increase in scores for older children than in younger ones. However, a univariate ANOVA with the two independent factors Age (three age levels: 8.6–8.11; 9.0–9.5; 9.6–9.11) and Gender (M, F) performed on the CPM raw scores of children from Sicily revealed no significant difference between age groups, $F(2,282) = .005, p > .05$, and genders, $F(1,282) = 1.69, p > .05$, nor interaction between components ($F(2,282) = 2.11, p > .05$).

This result is somewhat different from Cornoldi et al. (2010), who reported a difference in CPM scores between children of different age groups. It should be considered, however, that in Cornoldi et al.'s study the ages ranged from 7.6 to 9.11 years (for a total of 5 age groups), whereas in our study the age differences were more limited, ranging from 8.5 to 9.6. Moreover, Cornoldi et al. (2010) report in their study only the main effect of age, so that it is not possible to determine if there was a difference between the age groups that we have considered in our research.

Despite the minor differences between the studies, our results demonstrate quite clearly that raw scores of children from Sicily are not lower than those reported by Cornoldi et al. (2010). On the contrary, they are sometimes higher. This result could be related to the fact that the children in our group were tested in group sessions, while children in Italian standardization scores (Belacchi et al., 2008) were tested both in group and individual administration. Belacchi et al. (2008), indeed, found mean raw scores significantly higher in group sessions administration than in individual administration. Moreover, the children in our group were selected for other research purposes, and not included the children with socio-cultural disadvantage or other type of behavioral cognitive problems. The more extensive sample reported by Cornoldi et al. (2010), on the contrary, was collected with the aim of building norms, and it likely includes a more diverse sample of children coming from different urban and suburban areas, and showing different socio-cultural levels.

4. Study 2. Cognitive Assessment System's data

We also obtained data from a recently developed intelligence test that does not include typical verbal tests; for example, it doesn’t evaluate a child’s vocabulary in a specific language and it not even uses quantitative questions which require solving math word problems. Although the use of verbal and quantitative items is well entrenched in traditional tests for measuring IQ (Matarazzo, 1992), choosing a non-traditional intelligence test based on neuropsychological abilities makes a test more appropriate for culturally and linguistically diverse populations (Fagan, 2000).

Such an approach also provides a way to better understand children’s abilities and disabilities (Ceci, 2000) and is a more comprehensive way to view ability (Sternberg, 1988). More specifically, we utilized the PASS theory of intelligence measured by the Cognitive Assessment System (CAS; Naglieri & Das, 1997).

The PASS theory is based on the neuropsychological work of A. R. Luria (1973, 1980, 1982), and comprised four neurocognitive abilities Planning, Attention, Simultaneous, and Successive (PASS, Naglieri & Otero, 2011). The theory has undergone considerable evaluation. For example, studies of race and Ethnic have found only small differences between White and African-American (Naglieri, Rojahn, Mattio, & Aquilino, 2005) as well as between Hispanic and White children in the US (Naglieri, Rojahn, & Mattio, 2007). There are also research versions in Spanish (Naglieri, Otero, DeLauder, & Mattio, 2007) and Dutch which also show small differences across the US and the Netherlands (Van Luit, Kroesbergen, & Naglieri, 2005). These small differences suggest that ability may be more equitably assessed across race and

<table>
<thead>
<tr>
<th>Age</th>
<th>North* (Lombardy, Emilia Romagna, Friuli Venezia Giulia, and Veneto)</th>
<th>Central-south* (Abruzzo and Apulia)</th>
<th>South (Sicily)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.6–8.11</td>
<td>N 110</td>
<td>81</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>M 25.68</td>
<td>25.04</td>
<td>28.92</td>
</tr>
<tr>
<td></td>
<td>SD 5.39</td>
<td>5.90</td>
<td>5.23</td>
</tr>
<tr>
<td>9.00–9.5</td>
<td>N 57</td>
<td>109</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>M 27.4</td>
<td>26.35</td>
<td>29.18</td>
</tr>
<tr>
<td></td>
<td>SD 4.60</td>
<td>5.54</td>
<td>4.96</td>
</tr>
<tr>
<td>9.6–9.11</td>
<td>N 54</td>
<td>108</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>M 27.37</td>
<td>26.51</td>
<td>29.26</td>
</tr>
<tr>
<td></td>
<td>SD 4.71</td>
<td>4.64</td>
<td>5.03</td>
</tr>
</tbody>
</table>

* Data from Cornoldi et al. (2010).
Ethnic groups with a neuropsychologically based measure of ability. Importantly, the PASS abilities are highly correlated with achievement test scores ($r = .71, N = 1559$) for students aged 5–17 years (Naglieri & Rojahn, 2004). The most recent study (Naglieri et al., submitted for publication) involved Italian children, providing a way to evaluate differences across regions of that country.

4.1. Method

The Cognitive Assessment System — Italian Edition (CAS-I; Naglieri & Das, 2006) was administered to Italian children and adolescents in order to develop Italian norms (see Taddei & Naglieri, 2006 for more information).

4.2. Participants

The participants were 809 Italian subjects (aged 5, 6, 7, 9, 11, and 13 years) attending 34 schools distributed across Italy. The sample included students attending public schools from central (78.2%), northern (6.8%), and southern (15.0%) Italy who varied in socioeconomic characteristics, educational settings, and location of residence (e.g., central or peripheral areas of cities, urban or provincial settings). Similar numbers of males and females were obtained (overall 50.4% females and 49.6% males).

4.3. Materials and procedure

The Italian version of the CAS was individually administered by trained examiners. All procedures for standardized administration and scoring were closely followed. Raw scores on the 12 CAS subtests were converted to standard scores using the Italian normative values provided in the test manual.

4.4. Results and discussion

Naglieri et al. (submitted for publication) studied the differences between the psychometric properties of the CAS for the Italian and US standardization samples. Although the goal of that study was not to make regional comparisons, they did report that there were no significant differences ($F(1, 806) = 2.19, p = .11$) between the average CAS:Italian Full Scale standard scores (set at a mean of 100 and standard deviation of 15) for students from the northern ($M = 100.5, SD = 13.2$), central ($M = 101.2, SD = 11.9$), and southern ($M = 103.1, SD = 11.6$) regions of Italy. The mean standard scores for the students in the north were only slightly lower than the mean for those in the south (effect size = .21). These results suggest that a test of intelligence that measures basic neuropsychological processes, and does not include academically laden verbal and quantitative tests, yields small differences between the regional groups. These findings also amplify the importance of measuring intelligence directly when comparing groups and argue against using reading, math and science test scores as “proxies for intelligence” (Lynn, 2010a).

5. General conclusions

Our examination of intelligence test score differences between the north and south of Italy led to results that are very different from those reached by Lynn (2010a). Our results demonstrate that by using intelligence tests to assess differences in ability rather than using achievement scores as a proxy for intelligence, children from the south of Italy did not earn lower scores than those from the north of Italy. Rather, they were even higher in Raven’s CPM. However, we see no advantage in claiming that children in the south are “more intelligent” than children in the north, because these groups are different on a number of variables (e.g., environmental factors, educational influences, composition of the samples) that influence differences in test scores. We also disagree with Lynn’s genetically-centered explanation of intelligence which denounces a fixed conception not only about intelligence but also about learning. Carol Dweck (2000) and many others have amply demonstrated that such a conception of intelligence and learning has a destructive impact in students, teachers, and the educational system, and resulting problems such as learned helplessness and school drop-out.

We agree that intelligence is associated with scholastic achievement (Naglieri & Bornstein, 2003), and that the education–intelligence relationship is presumably reciprocal. As stated by Rindermann (2008, pp.138) “schooling raises intelligence, and intelligent people realize the advantages to be gained through a better education”. This does not mean that intelligence is scholastic achievement or vice versa (and therefore the latter can be used as a proxy to measure the former).

In conclusion, we are convinced that scientists should direct their efforts towards finding ways of reducing differences in educational attainment that may be due to differences in cultural and linguistic background. Moreover they should work at offering educational opportunities rather than use academic achievement scores as a “proxy for intelligence”.

References


A. D'Amico et al. / Learning and Individual Differences 22 (2012) 128–132


