British Journal of Educational Psychology (2005), 75, 339–350
© 2005 The British Psychological Society
The British Psychological Society
www.bpsjournals.co.uk

The equal environments assumption of classical twin studies may not hold

Ken Richardson[1] and Sarah Norgate[2*]
[1]Killin, UK
[2]The Open University, UK

The classical twin method – comprising comparisons of monozygotic (MZ) and dizygotic (DZ) twins – in the domain of cognitive abilities and attainments has led to wide acceptance of results suggesting a large amount of additive genetic variance, with far-reaching implications both for the nature of future studies on the causes of cognitive variance and for intervention policies, as in education. However, this interpretation is only valid if the method observes a number of conditions, which have to hold. Here, we show that the most crucial of these, namely, the equal environments assumption (EEA), may not hold. Consequently, differences in twin correlations might be at least partly explained by treatment effects from parents, teachers, peers, and so on. In addition, well-known interactions at various levels confound the model of simple additive effects on which the classical twin method is predicated and results are interpreted. For example, at a socio-cognitive level, DZ twins may respond to treatments differently from MZ twins. This interaction may further explain MZ–DZ correlation differences. There is abundant evidence for such interactive effects in published twin data. We suggest that there is a need for a more thorough examination of these problems.

The classical twin research design has been frequently used to describe sources of variance in cognitive abilities and attainments. Because monozygotic (MZ) twins are genetically identical, whereas dizygotic (DZ) pairs share only half their variable genes, on average, it is assumed that differences in intra-pair correlations can be used to estimate the amount of the total variation that is genetic in origin. The practice of doubling the difference to obtain estimates of heritability (the proportion of trait variation attributable to genetic variation within the population sampled) is now commonplace. Heritabilities of .5 – .8 for cognitive abilities and attainments have been widely accepted (Plomin & DeFries, 1998; Scarr, 1997; Walker, Petrill, Spinath, & Plomin, 2004). However, the conceptual simplicity underlying the method requires that a number of empirical preconditions be met. As Rutter (2002, p. 4) explains, 'all too often there has been a cavalier ignoring (. . .) of violation of key assumptions of the twin

*Correspondence should be addressed to Dr Sarah Norgate, Centre for Childhood, Development and Learning, The Open University, Walton Hall, Milton Keynes, Bucks, MK7 6AA, UK (e-mail: s.h.norgate@open.ac.uk).

DOI:10.1348/000709904X24690

design'. Although there are many examples of the possible violations (for a review, see Richardson, 1998), in the first part of this paper we focus on the most crucial of these, the equal environments assumption (EEA). The EEA presupposes that treatment effects, through which parents, teachers, peers and others treat the different kinds of twins differently, are randomized across the two sets of twins. In this paper, we review the evidence as to why the EEA may not hold for cognitive abilities and attainments. Where appropriate, we highlight associated educational and psychological implications.

Surprisingly, there have only been three attempts to test the EEA. Each of these attempts seems to have methodological problems. Also discussed is another key assumption of this design. The practice of estimating heritabilities from twin correlations (e.g. by doubling the MZ–DZ correlation difference) assumes that variance-creating genetic and/or environmental factors only have additive/independent effects. However, there is now considerable evidence for interactions between factors at, and between, several levels of development. At one extreme, interactions between genes, and between genes and environments, have been well demonstrated in animal research, such that there appear to be very few truly independent/additive gene factors underlying complex traits (Glazier, Nadeau, & Aitman, 2002). At the other extreme are socio-cognitive interactions. For example, because of the way they have been treated differently, DZ pairs may have different perceptions of their social worlds, and adapt to them as active centres of cognition, rather than as passive recipients of genetic and environmental forces, with different self-concepts, different learning aspirations, and so on, compared with MZ pairs. These interactive effects might further depress DZ correlations and further explain MZ–DZ correlation differences. In this study, we consider evidence that such interactive effects are present in published twin correlations.

Together, these problems can readily account for the classical patterns of twin correlations and compromise the twin method for genetic research, except in the very special circumstances that we raise further in this paper. It is noteworthy that this critique also applies to more recent statistical model-fitting analyses to twin data, which are based on similar assumptions. In the Discussion, we will also consider the implications of interactive, as opposed to additive, models of cognitive development and variation for social policy, especially education.

The EEA problem in MZ–DZ designs

The main problem with the classical twin design is captured by Kendler, Neale, Kessler, Heath, and Eaves (1993) as follows:

> The traditional twin method, as well as more recent biometrical models for twin analysis, are predicated on the equal environment assumption – that monozygotic (MZ) and dizygotic (DZ) twins are equally correlated for their exposure to environmental influences that are of etiological relevance to the trait under study.

Hettema, Neale, and Kendler (1995, p. 327) explain that, 'The validity of the EEA is crucial to these studies because, if incorrect, excess resemblance of MZ twins compared to DZ twins usually ascribed to genetic factors could be partly or entirely due to environmental factors'.

It is known that numerous aspects of home and other experience are much more similar for MZ twins than for DZ twins. In one recent review, Evans and Martin (2000, p. 77) state that,

> There is overwhelming evidence that MZ twins are treated more similarly than their DZ counterparts. As children, MZ twins are more likely to have the same playmates, share the same room and dress alike. As adults, MZ twins are more likely to keep in contact, than same-sex DZ twins.

In another review, Joseph (2000) cites questionnaire studies revealing very large differences between MZ and DZ pairs in experiences such as identity confusion (91% vs. 10%); being brought up as a unit (72% vs. 19%); being inseparable as children (73% vs. 19%); and having an extremely strong level of closeness (65% vs. 19%). It is also known that parents 'hold more similar expectations for their MZ than DZ twins with respect to social responsibility and independence' (Scarr & Carter-Saltzman, 1979, p. 528).

A proper test of the EEA cannot, of course, be straightforward because it is necessary to understand the environmental variables that might causally be involved in trait variance. These will almost certainly be different for different traits. As Sternberg (1995, p. 260) says with respect to IQ, '[psychologists] do not have a very good understanding of the factors that affect IQ. Neither does anyone else', Bouchard (1997) agrees that, 'in spite of years of concerted effort by psychologists, there is very little knowledge of the trait-relevant environments that influence IQ'. Much the same applies to other aspects of ability, such as school attainment. It is usually assumed that if MZ and DZ samples are balanced for a few demographic correlates of abilities or attainments, such as the socio-economic status of the home, then similarity of trait-relevant environments has been established and effects will be suitably randomized. However, this practice does not provide a genuine causal connection with trait variance, which might actually arise from other factors not balanced across twin types.

In spite of these uncertainties about the EEA, some claims are made in its defence. 'This possible confounding effect has been examined and, in research to date, does not appear to represent a major problem for the twin design' (Plomin & Daniels, 1987, p. 3). Walker *et al.* (2004, p. 329) claim that the EEA 'is supported by numerous studies'. In what is probably the strongest recent defence of the EEA, Bouchard (1997) cites 11 papers in support of the EEA, concluding that, 'A large number of studies have now been carried out on this problem' (p. 135). However, further examination shows that only three of the 11 papers are directly concerned with the EEA for cognitive abilities or attainments, with the remainder being about personality traits, recitations, or book reviews. Likewise, Walker *et al.* (2004), cite three studies, however, only one of them is related to cognitive abilities or attainments, with the others being about psychiatric illness. Indeed, across the numerous reports of twin studies of cognitive abilities/attainments of which we are aware, support for the EEA almost invariably relies on one or more of only three studies, all of which were conducted in the 1970s (those of Loehlin & Nichols, 1976; Matheny, Wilson, & Dolan, 1976; and Scarr & Carter-Saltzman, 1979). It is to an examination of these early studies that we now turn.

Defending the EEA

The study most widely cited in support of the EEA is that of Loehlin and Nichols (1976). They found little association between how much the twins played together, were dressed similarly and so on, and intra-pair differences on the National Merit Scholarship Test. It is the absence of such an association that was used to support the EEA.

This conclusion illustrates two problems common to all defences of the EEA to date. The first is that any strategy that searches for the absence of an association (or 'proving a negative') is acutely prone to Type II errors, which is increased because measures contain different sources of variance that should be controlled. Loehlin and Nichols also used mailed self-reporting techniques, which have weaknesses of biased impressions, poor memory, and poor reliability (McCartney, Harris, & Bernieri, 1990). For example, such measures typically show reliabilities of only 0.3–0.5 (e.g. Ge & Chadoret, 1996; O'Connor & Deater-Deckard, 1998).

This unreliability is compounded by a second problem, which was the use of difference scores as correlates. As Bouchard (1997, p. 134) states, 'Statistically sophisticated readers may note that difference scores are notoriously unreliable'. Loehlin and Nichols (1976) themselves say that their results do 'not altogether exclude a completely environmentalist position. The data do, it seems to us, have something important to say concerning the environment. And the upshot of what they say is that, it operates in remarkably mysterious ways' (p. 94). However, this does not convince us of the scientific rigour of the EEA.

The second study often cited is that of Matheny *et al.* (1976). They used 121 MZ twin pairs and 70 DZs that had appeared in a large longitudinal study of child development in the USA. The subjects had an age range of 3.5–13 years. Scores of physical similarity were obtained by sending a questionnaire to parents containing 10 items on physical resemblance (including facial appearance, height, whether or not the twins were ever confused for one another etc.). Intra-pair differences on the composite scores that resulted were then correlated with differences in intelligence test scores (the Stanford-Binet or the WISC depending on age). No significant association was found, and, support for the EEA was thus concluded.

However the unreliability of the study's difference measures casts doubt upon its conclusions. Also the small numbers of DZs in comparison to the numbers of MZ twins (70 vs. 121 pairs) suggests that parents were more likely to have responded if their twins were similar (which would reduce associations). In addition, there are problems stemming from the peculiar structure of the physical similarity scale. On six of the questions (e.g. height, hair colour, facial appearance etc.), parents were asked to respond with a score of 0 (*not at all similar*), 1 (*somewhat similar*), or 2 (*exactly similar*). With another four questions (e.g. 'is it hard for strangers to tell them apart?', 'Does father/mother ever confuse them?'), only a dichotomous scale was used: 0 (*no*) or 2 (*yes*). Responses were then added together to create a 20-point scale of twin similarity. Such similarity scores will produce acute restrictions of range. Indeed, 90% of MZs had similarity scores ranging from 12 to 20, while 80% of the DZs had scores ranging from 0 to 8. A rank order correlation was used, but with large numbers of pairs ranked on scores from 0 to 8 or 12 to 20, there are, inevitably, many tied ranks, which further reduces the possibility of significant association. In addition, the twins covered a wide age range, from 3 to 13 years, and although results are presented for younger and older children separately (age ranges not given) it is not clear that age ranges are homogeneous across zygosity groups.

It is noteworthy that such results report IQ-difference associations with parental perceptions of twin similarity, rather than how similarly parents actually treated the pairs, which is what a test of the EEA actually demands. However, Matheny *et al.* (1976) also asked parents to rate the frequency with which they dressed their twins alike, which can be taken as a rough index of actual treatment. There was a significant correlation between physical similarity and similarity of dress for the fraternal twins (but not for the MZ twins, which is to be expected given the restriction of range). Furthermore, the authors stated that 'for both identical and fraternal twins, there were significant correlations between socio-economic status and similarity of dress (. . .) the trend was for parents of lower socio-economic status to dress their children more alike' (p. 349). Since a correlation between socio-economic status and IQ is well known, this result suggests more intense similarity of treatment toward the lower range of IQ with possible attendant effects on IQ difference scores. Such an interaction may further detract from the overall correlation being sought. Again, we suggest that this study does not constitute a very powerful test of the EEA.

The third relevant study is that reported by Scarr and Carter-Saltzman (1979), which focused on disagreements among adolescent twins about their true zygosity. Although parents are rarely wrong about their twins' zygosity (e.g. Reitveld *et al.*, 2000), twins themselves seem less certain. In this sample of 174 pairs of MZs, only 104 agreed that they were MZ. Of 169 DZs, only 101 agreed they were DZ (i.e. 60% agreement; true zygosity assessed by blood typing). The point of the study was to examine (i) whether the IQ resemblances among pairs were more associated with their true zygosities than their perceived zygosities, and (ii) whether their facial and physical resemblance (as a presumptive index of 'treatment' by others) was associated with resemblance in IQ.

Twins' perceived zygosities and similarities of appearance were explored through questions such as, 'Do you and your twin look as alike as carbon copies?'. Other people's perceptions of twins' resemblances in appearance were estimated from ratings of black and white photographs of faces by eight independent psychologists. Various anthropological measures (height, weight etc.) were taken, as were differences in a dozen blood group markers. The children, who were between 10 and 16 years old, were tested on four different tests: the Raven's Matrices Test; the Peabody Picture Vocabulary Test; the Columbia Mental Maturity Scale, and the Revised Visual Retention Test (a memory test). In some tables, these scores are presented separately, and sometimes as an average cognitive measure, (though the reasons for the latter are unclear). With some comparisons, there was 'a highly inconsistent pattern of results across tests' (Scarr & Carter-Saltzman, 1979, p. 532).

According to Scarr and Carter-Saltzman, the first results show that, 'On intellectual measures co-twins were found to resemble each other according to their true, not self-perceived, zygosity' (p. 532; i.e. the data supported the EEA). The average absolute differences on the standardized scores are presented in Table 1 (taken from Scarr and Carter-Saltzman, 1979, Table 2). Apparently, these results are not as categorical as the authors suggest. For instance, those MZs who think that they are DZs (presumably, because they look or behave less alike) have inter-twin differences that are substantially greater than those who agree that they are MZs (first row of correlations). Indeed, they are only slightly smaller than those DZs who agreed they are in fact DZs (third column of correlations). Also problematic is the fact that only results using the pooled test scores are reported. This is of particular concern because the authors say that the differences fail to reach statistical significance because of 'a highly inconsistent pattern of results

Table 1. Standardized IQ score differences for MZ and DZ twins: Means of four tests (numbers of pairs in brackets). Source: Scarr and Carter-Saltzman (1979)

	Agree MZ	Disagree about zygosity	Agree DZ
True MZs	0.66 (89)	0.71 (41)	0.77 (20)
True DZs	0.78 (16)	0.83 (33)	0.81 (84)

Table 2. Correlations for general cognitive ability reported by Swedish Adoption/Twin Study of Ageing (SATSA). Source: Pederson *et al.* (1992)

	MZT	DZT	MZA	DZA
Full original sample	.80	.22	.78	.32
Longitudinal sample, first testing	.84	.06	.84	.50
Longitudinal sample, second testing	.88	.03	.70	.48

across tests' (Scarr & Carter-Saltzman, 1979, p. 532), and every other Table in the paper gives the scores of the different tests separately. Again, we stress that such correlational results provide no direct test of the EEA because self-perceptions of similarity, differences in cognition and treatment effects may be causes or results of each other.

Scarr and Carter-Saltzman's (1979) main results were from linear regression analyses in which intra-pair difference scores on IQ tests formed the dependent variable with all the other ratings and measures of similarity as independent variables. For the MZ twins, it was found that 'The Ravens Matrices score differences were significantly predicted by differences in appearance as rated by eight psychologists' (p. 535). This result confounds the EEA (on the assumption that differences in appearance lead to differences in treatment). The regression results for the DZs 'yielded a mixed picture' (p. 536), as they did for the combined sample, in that larger differences in Raven's scores were associated with smaller differences in appearance ratings and with larger differences in blood groups (the only other significant IV).

Again, these results do not offer unequivocal support for the EEA. In any case, there are many problems in interpreting such analyses. The coefficients that emerge estimate 'effects' for each variable when the effects of all the other variables have been accounted for. An important assumption is that the effects of all variables are independent of each other. The dozen or so similarity measures used by Scarr and Carter-Saltzman (1979) are clearly not independent of one another. In addition, since all of these factors accounted for less than 3% of the total variance in Raven's difference scores, there are obviously many variance explaining factors missing. Glymour (1997) has shown that omitted variables in regression studies can seriously distort coefficients. As Cohen and Cohen (1983, p. 92) explain, interpretation resides 'not in the coefficients but in one's understanding of the causal processes that give rise to the coefficients'.

The difficulty of such an understanding in the Scarr and Carter-Saltzman study is shown in the inconsistency, not only between test scores, but between all of the other ratings reported. Thus, twins' judgments of their own zygosity correlated only 0.55 with self-ratings of resemblance. Also surprising was that ratings of twins' zygosity by psychologists using only black and white photographs turned out to be more accurate than those of the twins themselves (correlation of .66 with blood group differences compared with only .57). The psychologists' ratings and twins' judgments of similarity also correlated only moderately at .51. Thus, twins' perceptions of their similarity are not very reliable guides to either their mutual resemblance, or to others' perceptions of their resemblance (and, hence, others' treatment effects). Again, we stress that finding an association in such unreliable data is highly unlikely even where one exists, and the idea that a genuine test of the EEA has been carried out by Scarr and Carter-Saltzman (1979) seems overly optimistic.

In sum, there are methodological problems in the few attempts to assess the EEA for cognitive abilities and attainments that have been conducted. We submit that these studies do not support the validity of the EEA. In the next section, we examine important aspects of twins' behaviour which, as well as offering further explanation for differences in twin correlations, serve to further undermine the EEA.

Interactions further explain patterns of correlations

Different treatment effects can explain at least part of the difference in correlations between MZ and DZ pairs. However, an alternative explanation is possible. The classical twin design, including the interpretation of correlations, is based on the assumption that cognitive variation reflects the sums of additive effects of genetic and environmental factors (i.e. with little, if any, effects of gene-environment interaction or of adaptive processes within the organism itself). Yet, an outpouring of research and theory at the molecular and epigenetic levels has increasingly revealed development through 'complex interactive systems with cascading levels of organization and numerous feedback systems' (Rollo, 1995, p. 6; for other reviews see Carroll, Grenier, & Wetherbee, 2001; Richardson, 1998).

Interactions at the genomic and epigenetic levels are further embedded in behavioural, cognitive, and (in humans) socio-cognitive interactions, which can further modulate simple additive effects and reduce predictability. For example, animal studies report how hormonal regulation of genes in response to stress can be transmitted from one generation to the next through maternal behaviour with offspring (Francis, Diorio, Liu, & Meaney, 1999). Research with children has shown how physical factors like height and facial appearance influence others' treatments in a way which can affect self-concepts and behavioural variation (see Richardson, 1998, for discussion). Scholastic attainment involves interactions between parental self-efficacy beliefs, aspirations for, and encouragement of, their children, and the latter's perceptions of the importance of educational attainments (e.g. Bandura, Caprara, & Pastorelli,1996).

Interactions at the socio-cognitive level may impact on twin correlations in several ways. By virtue of having cognitive systems and being part of social systems, MZ and DZ pairs will not only be treated differently, but also perceive, reason about, and react to treatments differently as individuals. DZs tend to be physically more distinct than MZs. In his review of research, Bryant (1992, p. 136) says that, 'It is inevitable that parents should constantly compare and contrast (. . .) and hence look out for differences (. . .; which) often leads to the exaggeration of character traits that in fact lie well within the normal range (. . .) parents and friends soon learn these stereotypes and, perhaps unconsciously, respond accordingly. More serious, however, is the fact that children tend to live up to parental expectations of being the one who is 'good' or 'naughty', 'quiet' or 'noisy''.

Future research needs to focus on the extent to which DZ's tend towards distinct identities in the domain of cognition. The development of polarized identities could shift motivations away from similar learning in literacy, numeracy, academic, and other ability-related domains. This is likely to be an active, not merely a passive, process, in which twins consciously shape different identities. In a review of parental ratings of temperament, Saudino (2003) has shown how such differences between DZTs begin early in infancy. Stewart (2000) describes how the social construction of identity in childhood can differ between MZ and DZ pairs, and across cultures. Akerman and Suurvee (2003) describe the special difficulties twins have in developing separate identities, and how this process varies between MZ and DZs. Schachter (1982) has hypothesized a process of de-identification among same-sex siblings as a result of competition or rivalry. It has been shown that DZs form individual identities earlier than MZs (Bryant, 1992). Segal (1984) found less cooperation among DZs than MZs. Loehlin and Nichols (1976) found that some twin pairs may attempt to differentiate themselves from each other through what they refer to as 'effects of competition or contrast' (p. 93). Thus, these authors reported a significantly greater similarity in the activities of MZs than DZs at 6 years. These distinguishing effects will also explain the increasing difference between MZ and DZ IQ correlations with age (McCartney *et al.*, 1990), which have usually been interpreted as a genetic effect on the basis of assumptions being challenged in this paper.

Such difference-creating interactions, reducing resemblances among DZs reared together (DZTs), are evident in published twin correlations. Indeed, Plomin and Loehlin (1989) alluded to contrast effects to explain puzzling results in twin correlations. The effects are perhaps best illustrated in the correlations of DZ twins who do not experience such rivalry (i.e. those pairs reared apart; DZAs). We would expect that, with the simple additive model, correlations would be substantially reduced from that of around .5 for DZs reared together. However, this does not seem to be the case (see Kamin & Goldberger, 2002; and Joseph, 2003 for review). For example, McGue and Bouchard (1989) reported a mean $r = .49$ for DZAs for the five subtests in a verbal reasoning factor derived from a special abilities battery. Newman, Tellegan, and Bouchard (1998) report $r = .47$ for 26 DZA pairs tested with the WISC. For a measure of verbal reasoning, the rDZA was .53. McCourt, Bouchard, Lykken, Tellegen, and Keyes (1999) reported a correlation of .53 for general cognitive ability for another subsample of 38 DZAs (which erroneously included 10 opposite-sex pairs; cf. Joseph, 2003). In a more general review, Plomin and Loehlin (1989) report an IQ correlation for DZAs of .52. The effect was also found in Shields' (1962) study in which the *r*DZA was .51.

More interesting are results from studies that made direct comparisons with *r*DZTs, such as the Swedish adoption/twin study of ageing (SATSA), which was also critically analysed by Kamin and Goldberger (2002). For example, with respect to the information scale of the Wechsler Test, Pedersen, Plomin, Nesselroade, and McClearn (1992) reported the following correlations: DZT = .20, DZA = .33. A measure of general cognitive ability (the first principal component from a battery of 13 brief subtests) produced the following correlations: MZT = .80, DZT = .22, MZA = .78, DZA = .32 (Pederson *et al.*, 1992). Another wave of testing 3 years later produced the correlations shown in Table 2 (Plomin *et al.*, 1994; cf. Kamin and Goldberger, 2002). It is unlikely that these anomalous correlation contrasts can be explained by extraneous factors such as age or gender differences. For example, SATSA researchers corrected results for age and gender effects, and multiple regression analysis showed 'no significant effects of age on twin resemblance' (Pederson *et al.*, 1992, p. 351).

The higher than expected correlations among DZ pairs reared apart (where contrast effects will be absent) suggests the depression of correlations among DZs reared together (where such effects are present). The correlations are difficult to explain within the standard behaviour genetic framework. They are more easily explained by reference to the socio-cognitive realities of twins; active cognitive systems immersed in social dynamics consisting of positive and negative socio-cognitive factors at work in complex interactions and in different proportions in different kinds of twins. Apart from explaining why *r*DZAs are as high as, or even higher than, *r*DZTs, these factors (at least partly) explain why *r*MZTs are higher than *r*DZTs. These social dynamics may also explain why heritabilities from direct estimates (biological relatives reared apart where the contrast interactions are absent) are much higher than those from indirect estimates (relatives and non-relatives reared together, where those interactions are present; Plomin & Loehlin, 1989). Finally, the inter-twin interactions also explain why DZs develop less resemblance as they get older, creating the illusion of rising heritabilities.

Discussion and conclusion

Many questions have been raised about the classical twin methodology over the years, mitigating interpretations about either genetic or environmental sources of variance. Nonetheless, the simple doubling of MZ–DZ correlation differences (or its parallel in model-fitting approaches) to obtain estimates of heritability of cognitive abilities and attainments has become increasingly common. In this paper, we have explored two problems that make such practice unreliable, indicating alternative explanations for MZ–DZ twin differences, and suggesting that the far-reaching conclusions often drawn from such data may be questionable. We have shown, first, that the EEA may not hold, and that well-demonstrated treatment effects can, therefore, explain part of the classic MZ–DZ differences. Using published correlations, we have also shown how socio-cognitive interactions, in which DZ twins strive for a relative 'apartness', could further depress DZ correlations, thereby possibly explaining another part of the differences. We conclude that further conclusions about genetic or environmental sources of variance from MZ–DZ twin data should include thorough attempts to validate the EEA with the hope that these interactions and their implications will be more thoroughly understood.

To promote the validation of the EEA there needs to be clearer understanding of the nature of the variance in cognitive abilities and attainments, and the nature of their environmental causes (which could then be properly controlled in EEA validation studies). The current assumption that equating across groups for a few physical and demographic correlates of trait variance amounts to controlling for real causal effects could be considered naïve in the light of the complex socio-cognitive and other interactions that may affect such variance. One 'short-cut' indicator of EEA validity may be to look at score variances of MZ and DZ twins, respectively. If these are equivalent, it may be argued that the EEA is valid (otherwise DZ variance would be significantly greater).[1] However, there are many reasons why equivalence of variance could not be taken as a test of the EEA's validity. For one thing, very few studies actually report test variances (see Kamin & Goldberger, 2002, for discussion). One reason for this may be that reported scores are often approximations or statistical reductions, which can severely affect variance components (e.g. the first principal component commonly used as a surrogate of *g* in recent twin studies typically accounts for only around 50% of raw score variance). In addition, sampling in twin studies is unsatisfactory in many ways that can alter relative score variances regardless of EEA. For example, relatively few same-sex DZ twins tend to be recruited in most studies, raising the suspicion that only parents of those pairs who are most alike physically respond to adverts. This may be exacerbated by poor socio-economic representativeness of samples (Rutter, 2002). In addition, it is well known that MZs tend to have more difficult gestations which, under the simple behaviour genetic model, could increase MZ variance (e.g. birth-weight variance is higher in MZs than DZs). On the other hand, most MZs are monochorionic and all DZs

---

[1] *We are grateful to an anonymous referee for raising this point.*

are dichorionic, which (under the same simple model) should increase DZ variance. Finally, it is found that not all MZ twins are actually genetically identical, and that DZ twins may share more than 50% of their genes (Gringras & Chen, 2001). The effects of these factors make expected MZ/DZ variance ratios quite uncertain.

Of course, it is now widely accepted that genetic and environmental variability may create variation in cognitive abilities and attainments. The important issue is the nature of that involvement. The classical twin design is predicated on the assumption that the involvement in both cases is a simple additive one in which the (variable) genes and (variable) environments act as independent units, allowing one-to-one predictability from variation in a given genetic/environmental factor to variation in cognition. However, reviews of the evolutionary genetics of complex traits in animals increasingly testify to the incorporation of independent alleles into interactive gene complexes, with attendant reductions in additive loci, and thus in heritabilities (in many traits to less than .1 or close to zero; see Merila & Sheldon,1999; Schonemann, 1997). Within genetics, it is now widely accepted that there are very few truly independent/additive gene factors underlying complex traits (Glazier, Nadeau, & Aitman, 2002). We would expect this picture to apply especially to highly evolved traits like human cognitive abilities.

Although often acknowledging interactions (e.g. Plomin, 1999), it seems strange that investigators keep on 'fitting' and favouring additive models, as if wishing to enjoy the force of argument both ways. However, the implications of the rival (additive vs. interactive) models for interventional policy, as in education, are serious. Multi-level interactive systems, which have evolved precisely for dealing with changeable, unpredictable environments, mean that it is very difficult to predict genuine developmental status from one time to the next, or the consequences of specific environmental experiences or interventions. This (among many other things) may explain why reported associations between single gene variations and variations in cognitive ability have failed to replicate (Plomin, 2003). They may also explain why raw correlations between school attainments and either university attainments or occupational attainments tend to be very weak (William, 2002).

Consequently, we suggest that standard behaviour genetic assumptions are only valid in specific cases (and at specific levels) where additive effects do indeed arise. These include the effects of well-known deleterious gene mutations, and teratogenic environmental effects, which produce categorical disease states, entailing a causal pattern quite different from that involved in 'normal' adaptive variation in cognitive abilities. Otherwise, the violation of the EEA in the classical twin design and ignoring interactions can lead to serious misinterpretations of results, as well as the controversies the studies have produced within psychology and elsewhere.

References

Akerman, B. A., & Suurvee, E. (2003). The cognitive and identity development of twins at 16 years of age: A follow-up study of 32 twin pairs. *Twin Research*, *6*, 328–333.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development*, *67*, 1206–1222.

Bouchard, T. (1997). IQ similarity in twins reared apart: Findings and response to criticisms. In R. J. Sternberg & E. Grigerenko (Eds.), *Intelligence, heredity and environment*. Cambridge: Cambridge University Press.

Bryant, E. (1992). *Twins and higher multiple births: A guide to their nature and nurture*. London: Hodder and Stoughton.

Carrol, S. B., Grenier, J. F., & Wetherbee, S. D. (2001). *From DNA to diversity: Molecular genetics and the evolution of animal design*. Oxford: Blackwell.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale: Erlbaum.

Evans, D. M., & Martin, N. G. (2000). The validity of twin studies. *GeneScreen, 1*, 77–79.

Francis, D., Diorio, J., Liu, D., & Meaney, M. J. (1999). Nongenetic transmission across generations of maternal behaviour and stress response in the rat. *Science, 286*, 1155–1157.

Ge, X., & Chadoret, R. J. (1996). The developmental interface between nature and nurture: a mutual influence model. *Developmental Psychology, 32*, 524–589.

Glazier, A. M., Nadeau, J. H., & Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science, 298*, 2345.

Glymour, C. (1997). Social statistics and genuine inquiry: Reflections on The Bell Curve. In B. Devlin, S. E. Feinberg, D. P. Resnick, & K. Roeder (Eds.), *Intelligence, genes and success*. New York: Copernicus.

Gringras, P., & Chen, W. (2001). Mechanisms for differences in monozygotic twins. *Early Human Development, 64*, 105–117.

Hettema, J. M., Neale, M. C., & Kendler, K. S. (1995). Physical similarity and the equal-environment assumption in twin studies of psychiatric disorders. *Behavior Genetics, 25*, 327–335.

Joseph, J. (2000). Not in their genes: a critical review of the genetics of attention-deficit-hyperactivity disorder. *Developmental Review, 20*, 339–367.

Joseph, J. (2003). *The gene illusion: Genetic research in psychiatry and psychology under the microscope*. Ross-on-Wye: PCCS Books.

Kamin, L.J., Goldberger, A.S. (2002). Twin studies in behavioural research: A sceptical view. *Theoretical Population Biology, 61*, 83–95.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics, 23*, 21–27.

Loehlin, J. C., & Nichols, R. C. (1976). *Heredity, environment and personality*. Austin: University of Texas Press.

Matheny, A. P., Wilson, R. S., & Dolan, A. B. (1976). Relations between twins' similarity of appearance and behavioral similarity: testing an assumption. *Behavior Genetics, 6*, 343–351.

McCartney, K., Harris, M. J., & Berbieri, F. (1990). Growing up and growing apart. *Psychological Bulletin, 107*, 226–237.

McCourt, K., Bouchard, T. J., Lykken, D. T., Tellegen, A., & Keyes, M. (1999). Authoritarianism revisited: Genetic and environmental influences examined in twins reared apart and together. *Personality and Individual Differences, 27*, 985–1014.

McGue, M., & Bouchard, T. J. (1989). Genetic and environmental determinants of information processing and special mental abilities: A twin analysis. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 7–45). Hillsdale: Erlbaum.

Merila, J., & Sheldon, B. C. (1999). Genetic architecture of fitness and nonfitness traits. *Heredity, 83*, 103–109.

Newman, D. L., Tellegen, A., & Bouchard, T. J. (1998). Individual differences in adult ego development: Sources of influence in twins reared apart. *Journal of Personality and Social Psychology, 74*, 985–995.

O'Connor, T. G., & Deater-Deckard, K. (1998). Genotype-environment correlations in late childhood and early adolescence. *Developmental Psychology, 34*, 970–982.

Pedersen, N. L., Plomin, R., Nesselroade, J. T., & McClearn, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the lifespan. *Psychological Science, 3*, 346–353.

Plomin, R. (1999). Genetics and general cognitive ability. *Nature, 402*, C25–C29.

Plomin, R. (2003). Genetics, genes, genomics and g. *Journal of Molecular Psychiatry, 8*, 1–5.

Plomin, R., & Daniels, D. (1987). Why are children in the same family so different from one another. *Behavioral and Brain Sciences, 10*, 1–60.

Plomin, R., & DeFries, J. C. (1998). The genetics of cognitive abilities and disabilities. *Scientific American*, *May*, 62–69.

Plomin, R., & Loehlin, J. C. (1989). Direct and indirect heritability estimates: A puzzle. *Behavior Genetics*, *19*, 331–342.

Plomin, R., Pedersen, N. L., Lichtenstein, P., & McClearn, G. E. (1994). Variability and stability in cognitive abilities are largely genetic later in life. *Behavior Genetics*, *24*, 207–215.

Reitveld, M. J. H., van der Valk, J. C., Bongers, L. L., Stroet, T. M., Slagboom, P. E., & Boomsma, D. I. (2000). Zygosity diagnosis in young twins by parental report. *Twin Research*, *3*, 134–141.

Richardson, K. (1998). *Origins of human potential: Development, evolution and psychology.* London: Routledge.

Rollo, D. C. (1995). *Phenotypes: Their epigenetics, ecology and evolution*. London: Chapman and Hall.

Rutter, M. (2002). Nature, nurture and development: From evangelism through science toward policy and practice. *Child Development*, *73*, 1–21.

Saudino, K. J. (2003). Parental ratings of infant temperament: Lessons from twin studies. *Infant Behavior and Development*, *26*, 100–107.

Scarr, S., & Carter-Saltzman, L. (1979). Twin method: Defense of a critical assumption. *Behavior Genetics*, *9*, 527–542.

Schachter, F. F. (1982). Sibling deidentification and split-parent identification: A family tetrad. In M. E. Lamb & B. Sutton-Smith (Eds.), *Sibling relationships: Their nature and significance across the lifespan* (pp. 123–151). Hillsdale, NJ: Erlbaum.

Schonemann, P. H. (1997). On models and muddles of heritability. *Genetica*, *99*, 97–108.

Segal, N. L. (1984). Cooperation, competition and altruism within twin sets: A reappraisal. *Ethology and Sociobiology*, *5*, 163–177.

Shields, J. (1962). *Monozygotic twins brought up apart and brought up together.* London: Oxford University Press.

Sternberg, R. J. (1995a). For whom the bell curve tolls: A review of the bell curve. *Psychological Science*, *5*, 257–261.

Stewart, E. (2000). A Towards the social analysis of twinship. *British Journal of Sociology*, *51*, 719–737.

Walker, S., Petrill, S., Spinath, F. M., & Plomin, R. (2004). Nature, nurture and academic achievement. *British Journal of Educational Psychology*, *74*, 323–342.

William, D. (2002). *Level best: Levels of attainment in national curriculum assessment*. London: ATL.

Received 4 July 2003; revised version received 1 March 2004