
Article:
Asbury, Kathryn orcid.org/0000-0003-0011-457X, Moran, Nicola Elizabeth orcid.org/0000-0002-6219-0394 and Plomin, Robert (2016) Nonshared Environmental Influences on Academic Achievement at Age 16:A Qualitative Hypothesis-Generating Monozygotic-Twin Differences Study. AERA Open. pp. 1-12. ISSN 2332-8584

https://doi.org/10.1177/2332858416673596

AERA Open
October-December 2016, Vol. 2, No. 4, pp. 1–12
DOI: 10.1177/2332858416673596

Nonshared Environmental Influences on Academic Achievement at Age 16: A Qualitative Hypothesis-Generating Monozygotic-Twin Differences Study

Kathryn Asbury
Nicola Moran
University of York
Robert Plomin
King's College London

Twin studies find ~20% of the variance in achievement in public examinations taken at age 16 in the United Kingdom can be explained by experiences not shared within families. Nonshared environmental (NSE) influences, including measurement error, explain why monozygotic (MZ) twins differ from each other. Such influences work independently of genetic effects and may represent strong candidates for intervention. This study aimed to generate hypotheses about what these NSE factors might be. Perceptions of within-pair differences were gathered from n = 497 pairs of MZ twins and their parents, and telephone interviews were conducted with n = 56 families reporting different General Certificate of Secondary Education grades. "Environmental" explanations related to teacher quality, teacher–pupil relationships, and ability grouping. Families also explained discordance in terms of effort, interest, ability, and personality.

Keywords: *twin study, behavioral genetics, nonshared environment*

We know that genes influence but do not determine academic achievement (Asbury & Plomin, 2013). One of the most powerful illustrations we have of this fact is that identical (monozygotic [MZ]) twins do not always achieve the same grades, although their achievement is more similar than that of less genetically related individuals. In some cases, MZ twins' academic performance is strikingly different, and the reason for this is likely to lie somewhere in the environments in which they were raised and educated or in their pre- or perinatal experiences. Understanding influential experiences matters a great deal in adolescence, a time when young people make important choices and when successes and failures can have long-term consequences. We know that MZ discordance cannot be genetic because MZ twins share identical genotypes, albeit with a miniscule chance of mutation. However, pinning down precisely which aspects of experience lead to discordance has proved not unlike hunting the proverbial needle in a haystack.

Behavioral geneticists partition environmental influences into those that are *shared* (contribute to the similarity of siblings brought up in the same family) and those that are *nonshared* (do not contribute to sibling similarity). Because most MZ twins share a home as well as their genes, differences between them can be explained only by experiences they do not share or that affect them differently (nonshared environment [NSE]). This study therefore involved an in-depth search through the haystack in order to identify candidate NSE influences on academic achievement. This is an important line of inquiry for educational research because if we can identify experiences that influence behavior independently of genes, they may represent strong targets for carefully designed interventions.

In the United Kingdom (apart from Scotland), all pupils take GCSEs (General Certificate of Secondary Education) or equivalent qualifications at age 16. Recent research has shown that GCSE performance is heritable: Differences between 16-year-olds in how they perform are influenced by differences in their genes (Krapohl et al., 2014; Shakeshaft et al., 2013). In one such study, Shakeshaft et al. (2013) gathered GCSE grades from 11,117 twins and found that 50% to 60% of the variance in core GCSE subjects (English, math, and science) was explained by genetic variance, 20% to 30% by shared environmental influences, and the remainder, approximately 20%, by NSE influences (including measurement error). Similar patterns have been observed at different ages and in different countries (e.g., Calvin et al., 2012). Researchers are working to identify both the genes that can explain these heritability estimates and the experiences that can explain

the environmental variance, but thus far, these endeavors have experienced limited success. We have both a "missing heritability" and a "missing environment" problem, although recent developments in molecular genetics give cause for cautious optimism that the genes that explain individual differences in cognitive ability will gradually be found by genome-wide association studies with very large samples (e.g., Okbay et al., 2016). The current study aimed to generate new, testable hypotheses about specific measurable NSE influences, with an eye to intervention. The challenge remains to take a similar approach to the shared environment and to genotype–environment interplay.

Identifying specific NSE factors that can explain variance in behavioral outcomes has proved more difficult than anyone imagined. There are some who believe the hunt is destined to failure because effects are likely to be too small or unsystematic to detect (Turkheimer & Waldron, 2000). These authors suggest that most NSE experiences are likely to be little more than chance occurrences. Their argument is based on a detailed meta-analysis of studies, including the Non-shared Environment and Adolescent Development study (Reiss et al., 1995), which found that measured NSE variables explained a negligible amount of NSE variance. However, it has also been argued that small NSE effects may accumulate to explain larger proportions of variance (e.g., Plomin & Daniels, 1987). Proponents of both arguments agree that although NSE influence is often substantial, we have not yet understood the causal mechanisms involved. Turkheimer and Waldron (2000) make the important point that objectively nonshared experiences are unlikely to be the only, or even the most important, contributors to NSE variance and that shared experiences can have nonshared effects. For this reason, it is essential to include subjective experience in studies designed to understand NSE influence. Studies should also expect small effects and take genetic effects into account (Turkheimer & Waldron, 2000). Furthermore, they should attempt to test whether candidate NSE experiences show causal relationships with behavioral outcomes and whether variance in purported NSE variables is substantially nonshared. The current study has been designed to identify potential NSE experiences and represents the first step in a two-step program of research. This hypothesis-generating study will be followed by a quantitative study that will assess the relationship between hypothesized NSE experiences and achievement, calculate whether variance in NSE experiences are nonshared in origin, and expect small effects.

Another possibility to consider is that NSE influences may not be stable over time, so they may lack predictive validity. Indeed, we know that genetic sources of variance for cognitive ability are more stable than nonshared environmental sources of variance (Tucker-Drob & Briley, 2014). This is to be expected as a growing child's experiences and environments can change enormously over time. What is more surprising is that this meta-analysis found that some NSE influence is not transient and some effects do indeed persist over time, with stability increasing in adolescence and adulthood. This suggests that NSE influences can have lasting effects and supports the proposal that NSE represents more than just measurement error. To illustrate the point, it is possible that if a measured NSE experience influences academic achievement, even if it seems like a transient experience, it may have lasting effects (e.g., a pupil may not get a high enough grade to be allowed to study a subject at the next level).

We believe there is value in looking for tangible explanations of NSE variance, not least because explaining even a small proportion of the differences between pupils in their academic achievement could pave the way for new and effective teaching and learning strategies (Plomin, 2011). Explaining NSE variance in GCSE is likely to be valuable even if relevant experiences do not generalize to individuals and families in countries and contexts where GCSEs are not taken. If any hypotheses explain variance in U.K. academic achievement at 16, then this will have a localized value. Despite differences, all researchers in the field agree that studying the differences between MZ twins is a sharp and effective route to understanding NSE. Designs that study similarities and differences in a variety of sibling types also have great value in this endeavor (e.g., Iervolino et al., 2002).

A body of research focusing on the causes and correlates of MZ discordance in a range of phenotypes was generated by the publication of a seminal review (Plomin & Daniels, 1987) that was described by Turkheimer and Waldron (2000) as "what may have been the most influential article ever written in the field of developmental behavioural genetics" (p. 78). The body of work inspired by this paper has identified NSE experiences associated with MZ differences in a wide range of outcomes in both cross-sectional and longitudinal designs. The majority of this research has focused on discordant parenting (e.g., Asbury, Dunn, Pike, & Plomin, 2003; Asbury, Dunn, & Plomin, 2006a; Caspi et al., 2004; Pike, Reiss, Hetherington, & Plomin, 1996; Burt, McGue, Iacono, & Krueger, 2006; Viding, Fontaine, Oliver, & Plomin, 2009). In these studies, MZ differences in parenting were found to correlate, in expected directions, with MZ differences in behavior problems and achievement. A pattern of small effect sizes (~3% or less) that sometimes get larger at the extremes (~10%) can be observed.

A handful of studies have also looked for NSE in the school environment. For instance, 61 pairs of 10-year-old MZ twins, each pair in the same class, were interviewed every day for 2 weeks in an attempt to identify NSE stressors in the school environment (Asbury, Almeida, Hibel, Harlaar, & Plomin, 2008). MZ differences in perceived peer, academic, and teacher stressors were modestly but significantly associated with MZ differences in "flow" in lessons (Csikszentmihalyi & Csikszentmihalyi, 1988) and positivity about school. For example, when one identical twin reported

more peer stress, he or she was significantly less likely than the co-twin to report being happy, engaged, or "in flow" during English lessons. MZ differences in peer stress also correlated significantly and in the expected direction with MZ differences in math achievement, suggesting that peer relationships may explain some of the variance in mathematics achievement at this age.

In another school-focused MZ differences study, data were gathered from 285 nine-year-old MZ twins, this time in different classrooms (Oliver, Pike, & Plomin, 2008). Researchers found substantial dissimilarity in individuals' experiences of the classroom environment (average $r = .40$). They also found that MZ differences in classroom experiences could account for 1% to 5% of the variance in teacher ratings of behavioral strengths and difficulties.

More recently, researchers asked whether NSE factors in kindergarten were associated with MZ differences in academic achievement in first grade (Vitaro, Boivin, Brendgen, Girard, & Dionne, 2012). In this case, MZ differences in peer rejection and teacher–child relationships in kindergarten were significant correlates of MZ differences in academic achievement in Grade 1, suggesting some predictive validity.

These studies suggest that experiences such as parenting and peer rejection are important and, more specifically, that they may act as NSE influences. If it is found that within-family experiences can explain between-family variance, these studies suggest that small effects should be expected. After identifying influential experiences, we will need to understand the interplay between each experience, other aspects of the environment, and individual genomes. For now, though, the challenge to identify candidate NSE factors remains, and such factors are needed for the development of new hypotheses and evidence-based NSE interventions for education.

There are several reasons why identifying NSE factors has been difficult. One is the sheer complexity involved, given that human behavioral traits are influenced by intricate and dynamic relationships between many aspects of both person and environment. Another is that we may not be measuring the environment accurately enough. Our measures of parenting and the classroom environment may not actually reflect pupils' experiences. This is a driving force for the current study in which we try to understand how young people and their parents really perceive the learning environment and which aspects of it they believe make a tangible difference to GCSE performance. Although student and parent beliefs do not necessarily represent the ways in which experiences actually influence GCSE performance, this perspective represents a strong starting place for a hypothesis-generating exploration. A small number of behavioral genetic studies have already taken the unusual (for this branch of psychology) step of taking a qualitative approach to develop a more nuanced understanding of NSE experience.

Most recently, a study of adult MZ twins, discordant for major depression, used autobiographical interviews to generate hypotheses relating to NSE influences on depression (Kendler & Halberstadt, 2013). Differences in the perceived quality of intimate relationships emerged as the most oft-mentioned NSE influence. We are also aware of three qualitative MZ twin studies of childhood experiences. In one, 62 primary caregivers were videotaped interacting with their three-and-a-half-year-old twins and were subsequently interviewed about their disciplinary strategies. Differences in observed parenting were found to correlate with differences in social-emotional adjustment (Deater-Deckard et al., 2001). In another study, expressed emotion was measured in mothers of 565 five-year-old MZ twin pairs (Caspi et al., 2004). Mothers were asked to talk freely about each twin in turn while being audio recorded. Twins about whom mothers expressed more negative emotion and less warmth were found to show more antisocial behavior problems, a clear indication of a correlational NSE relationship between parental warmth and child behavior. Another study used telephone interviews to explore why some identical twins are more anxious than their co-twins (Asbury, Dunn, & Plomin, 2006b). Mothers described discordance in negative school experiences, illnesses and accidents, neonatal life events, parent–child relationships, and peer rejection as explanations for discordant anxiety. The qualitative approach taken by these studies offers promise for gaining a more fine-grained understanding of the nonshared experiences of identical twins. We therefore used a qualitative MZ-twin differences design to generate hypothetical explanations for within-pair discordance in GCSE achievement. Our aims were to work toward developing precise measures of learning environments experienced by young people preparing to complete their compulsory education in the United Kingdom and toward a deeper understanding of which of those experiences might influence exam performance, independently of genes. The ideas offered by twins and their parents will be empirically tested in future studies and, if appropriate, used to inform evidence-based interventions.

Methods

Participants

Questionnaire data were gathered from $n = 497$ families with identical twins (61% female), from the U.K. Twins' Early Development Study (TEDS), a longitudinal study of twins born in the United Kingdom between 1994 and 1996 (Oliver & Plomin, 2007). Zygosity was confirmed using DNA for 84% (questionnaire) and 85% (interview) of participants. In the remaining cases, zygosity was assigned via a questionnaire that has been found to be 95% accurate in the TEDS sample (Price et al., 2000). The TEDS sample has been shown to be reasonably representative of the U.K. population of same-age adolescents and their parents (Haworth, Davis, & Plomin, 2013) but was not fully representative in the current study. We invited 2,162 TEDS families to take part, and of those, we received data from 497, a response rate of 23%.

This was lower than hoped, which may reflect sample selectivity. The relatively increased proportion of girls in the current sample (from ~50% at first contact to 61%) is representative of TEDS at 16 although not of wider U.K. society. This significant discrepancy may be the result of greater willingness to engage with data collection among girls than among boys at this age and stage. The current sample was also significantly higher in terms of socioeconomic status (M= 0.31, compared to 0.00 at first contact and 0.1 at age 16) and g (measured at age 12; M= 0.11, compared to 0.00). All group mean differences were assessed with t tests. TEDS families have been studied throughout their lives, but this was the first occasion on which we had asked them to provide free-response data. There are indications that the approach was off-putting to some, potentially leading to a slightly biased sample. Although this does not matter in one sense, because our interest was in within-pair rather than between-family differences, it is important to bear the evidence of sample selectivity in mind. It remains possible that NSE influences are different for families in different circumstances.

Three questionnaires were posted to each family, and in most cases, we received self-report data from a parent (usually mother) and both twins. The twins' average age was 17.3 (range = 16.2–18.9).

After analysis of the questionnaires, telephone interviews were conducted with 56 families in which pairs were at least two grades apart in at least one GCSE subject. We observed GCSE grade differences of two or more grades in 65 families and were successful in organizing interviews with 56 of them. In the remaining nine cases, we were unable to make contact, consent to participate was not given, or consent was given but the phone was not answered subsequently. In 51 families, both twins and one parent (usually mother) were interviewed; in three, just one twin and no parent was interviewed; in one family, both twins but no parent were interviewed; and in the final family, a parent and one twin were interviewed (n = 160 individuals).

Self-reported GCSE data had previously been collected shortly after the official release of U.K. school examination results in August 2010, 2011, and 2012. In England and Wales, GCSE examinations are taken at age 16 and are graded from A* (A-star) to G. We have previously shown that the self-reported exam results are reliable by verifying grades against the U.K. National Pupil Database, using a sample of 7,367 twins, yielding a correlation of 0.99 for mathematics, 0.98 for English, and >0.95 for science subjects (Rimfeld, Kovas, Dale, & Plomin, 2015).

Measures

Questionnaires. A screening questionnaire was designed to identify potential sources of discordance between identical twins toward the end of compulsory education. The first item asked whether twins performed differently in their GCSEs overall and, if so, what the differences were and how they might be explained. The second focused on discordance in core GCSE subjects—English, math, and science—and asked whether there was a difference of at least two grades (e.g., A*/B or D/F) and how such discordance might be explained. Items were open-ended as the aim was to ask families for their hypotheses in a way that would not be leading.

Interviews. Telephone interviews were conducted by two experienced interviewers. Because of the hypothesis-generating nature of this study, different interview guides were drawn up by the researchers for each participant, focusing on differences and explanations identified in the questionnaire. Researchers read the completed free-response questionnaires provided by each family in which twins differed by at least two grades in at least one GCSE subject. They then documented all reasons offered by each member of the family to explain this discordance and turned the explanations into questions followed by a series of relevant probes. This formed a semistructured interview schedule that differed by family. Also, when potential hypotheses were suggested in the interviews that had not been mentioned previously, interviewers probed for a full account of each participant's view. This flexible approach was taken so that participants could give a full account of their beliefs about why one twin performed better than the other, unrestricted by closed or standardized questions.

Procedure

Families invited to participate in the study received an information letter, consent form, and three questionnaires, one for a parent and two for the twins. Separate envelopes for each participant were included so that individuals would be able to keep their responses private. Families returning completed sets of questionnaires received a £15 voucher. On receipt, questionnaire data were transcribed and entered into Excel.

Analysis of questionnaire data served two related purposes: (a) to indicate possible explanatory factors for differences in achievement at GCSE between identical twins and (b) to aid selection of a subsample of families to be contacted for follow-up interviews.

Families selected for interview were contacted by telephone and were asked for consent to participate. Times were then arranged to interview all three family members participating in the study. In cases where all family members were interviewed during the same telephone call, they were asked not to be in the same room to ensure individual privacy. All interviews were recorded and transcribed with the full consent of participants.

Analysis

Explanations offered for discordance in GCSE achievement in the screening questionnaires were tallied and collapsed into broad themes. Difficulties were encountered

where twins took equivalent but different qualifications, for example, science qualifications organized and assessed by different examination boards. However, an A, or an E, in one GCSE qualification should be equivalent to an A, or an E, in any other. Therefore, stated grades were accepted and included if they were at least two grades apart. Explanations were documented in participants' own words, staying close to the original data, and were later categorized as being "about," for instance, effort or teachers. Most analysis was undertaken by a single researcher, but interrater reliability checks were carried out by another member of the team. Ten percent of the data set was checked in this way. The second researcher noted their own interpretations of possible explanatory factors, and these were subsequently checked against those recorded by the first researcher. There was a very high degree of agreement between the two researchers (89% agreement in a subsample of 50 families). Subtle differences were discussed and agreed upon, following which slight changes were made to the coding frame and categorization of potential explanatory factors.

All interview transcripts were fully anonymized and charted using the framework approach (Ritchie and Spencer, 1994) to order and synthesize data through five stages: familiarization, identifying conceptual themes, indexing, charting, and mapping. The framework approach allows the sequential organization and interpretation of qualitative data. A table is created that displays cases in rows and themes in columns. Taken together, the rows and columns suggest explanations. Interrater reliability checks were conducted on 10% of the interview data set, with two researchers reaching an equivalent degree of congruence to that achieved with the questionnaire data.

Results

Explanatory Factors Identified via Questionnaire

Sixty-five families reported differences of at least two grades in one or more core GCSE subjects in their questionnaires (see Table 1).

In total, 30 sets of MZ twins showed a two-grade difference in English GCSE, 23 in math, and 31 in science. A further 15 families reported a difference in grades but did not state actual grades. Because it could not be assumed that they were reporting differences of at least two grades, we included families only where we were confident. This was because discordance of a single grade could reflect very minimal discordance in performance, as little as 1% at the grade boundary.

Between them, these 65 families reported 101 possible explanations for differences in attainment in the three core GCSE subjects. Explanations reported by at least three different families in any one subject are summarized in Table 2.

Differential motivation or effort was the most frequently mentioned explanations for discordance across GCSE subjects, closely followed by ability and interest. These behavioral differences require further explanation at the environmental level.

TABLE 1
Breakdown of Grade Differences of at Least Two Grades, by Subject

Differences more than one grade apart	*n*
English only	22
Math only	11
Science only	16
English and math only	1
English and science only	4
Math and science only	8
All three core subjects	3
Total	65

Explanatory Factors Identified via Interview

Fifty-six of the families in which twins were discordant by at least two grades were interviewed. Interviews covered explanations offered in the questionnaires, which spanned the entire period from the mother's pregnancy through birth and neonatal experiences, preschool years, experiences of primary school, transition to secondary school, and GCSEs. All of these data were taken into account in considering potential influences—both direct and indirect—on MZ discordance in GCSE achievement.

It is important to reiterate that these data represent a series of case studies and cannot speak to direction of effects. Furthermore, a within-family effect does not necessarily mean there will be a related between-family effect. It is possible that findings from these case studies will relate only to the families themselves. This study was designed to identify within-family effects, potential NSE factors that may prove to be useful targets for intervention, but assessing whether this is the case will involve further empirical testing. Two key themes were identified in the analysis of interview transcripts:

- School environment
  - Ability grouping
  - Teacher quality
  - Teacher–pupil relationships
- Individual traits/behavior
  - Ability
  - Personality
  - Effort and motivation
  - Interest or enjoyment

School Environment. Of the 56 families interviewed, members of 42 perceived ability grouping, quality of teaching, and/or the teacher–pupil relationship as part of the explanation.

TABLE 2
Explanations Found in Questionnaire Data for Two-Grade Discordance in English, Math, and Science

Explanatory category	Subcategory	English	Math	Science	Total
Teachers	One had "better" teacher	3	2	—	5
	Different teachers/teaching styles in same subject	5	3	2	10
Ability grouping	Different sets	1	5	2	8
Personality	One more focused/determined/motivated	4	1	2	7
	Different people/individuals	2	3	1	6
	One finds it harder to concentrate	3	5	4	12
Ability	One understands more/better comprehension/finds subject easier	8	15	8	31
	One more academic/scientific/creative	—	—	3	3
Effort	One worked harder/put in more effort during GCSE period	6	7	7	20
	One revised more/harder for assessments	7	5	10	22
Interest	One more interested in the subject	8	5	9	22

Note. GCSE = General Certificate of Secondary Education.

Ability grouping. The general consensus among twins and parents was that students in higher sets tended to receive a better quality of teaching. This was variously described as the teacher explaining issues or concepts better, engaging more with students, having more passion for the subject, pushing students to reach their potential, or being better able to control the class.

A clear example of this was presented by one family in which twins were placed in different mathematics sets to each other, neither of them the top set. The twin placed in the higher set noted, "I was in a class where like everyone wanted to try and get a good grade." This twin reported that the teacher was influential and would tell students to knuckle down in order to get a good exam grade. In comparison, her co-twin, placed in a lower math set, felt that the focus was on higher achieving students at the expense of others.

> I wanted help but . . . it was all about the clever people. . . . I gave up towards the end because I knew I wasn't going to get the grade anyway.

This disenchanted student was awarded an F grade while her co-twin achieved a D. In this case, although neither twin was in a "top" set, there was still a perceived difference in teaching quality, expectations, and attitude between a higher and a lower set.

In a different family, with twins also placed in different sets for mathematics, the twin in the lower set reported low-level disruption among the more-able pupils in her set—students she perceived as finding their work too easy. The teacher was reportedly not good at dealing with this disruption.

> I think that just made a massive difference to the class. I know everyone saw them and thought, well, they are not concentrating so why should we?

Subsequently, despite a predicted B grade in math, this twin achieved a C while her co-twin achieved a grade A.

However, it should also be noted that there were exceptions, with some twins reporting that teachers of lower sets worked extra hard to help students succeed. For example, one twin in a lower set for English than his co-twin said of his teacher,

> She knew what students needed and she treated each one individually not as a group. . . . Everyone liked her because she really connected to the students and made it so much easier for them in any way that she could.

Although this teacher was appreciated by this twin, who reported having less encouraging and supportive English teachers in the past, he still performed less well (grade D) than his co-twin in a higher set (grade B).

Finally, ability grouping was perceived by twins and their parents as having an impact on motivation. One twin reported being moved up a set in math after working hard, and feeling motivated by this to work even harder. His co-twin said that his brother was driven to succeed because he had initially been placed in a lower set than him and simply did not accept that he was less good at math. He had something to prove, and the family believes that this explains the boys' eventual discordance in GCSE Physics (A vs. C). They reported that the twin who was originally placed in a higher set lost confidence as his brother became increasingly motivated and sped past him. It should be noted that the twin relationship dynamic should be considered in interpretation of these findings.

Perceived teacher quality. Parents and twins offered explanations for GCSE discordance relating to their perceptions of teacher quality in a variety of ways: inspirational teachers, absent/supply teachers, innovative teaching methods, and behavior management.

A number of families spoke of "inspirational" teachers. Passionate, enthusiastic teachers who were perceived as committed to helping all students achieve to the best of their ability were appreciated. One twin said,

> He was passionate about the subject; he was quite funny, he had a laugh and stuff when he was teaching it. But also the fact that he was very passionate about it and he wanted to pass that over to the students . . . .

This twin's English GCSE result was two grades higher than that of her co-twin taught by a different teacher. This family viewed the achievement discordance as a direct result of one twin being taught by a more effective teacher.

In some cases, there was a perceived lack of teaching. There were examples of teachers being absent for most or part of an academic year and students being taught by supply/cover teachers with little perceived interest in teaching them. In one family, one twin had been predicted an A in English, but as the teacher was not present for most of the year, the twin (and allegedly most of the class) got a D while his co-twin achieved a B.

Innovative teaching methods were valued. For example, one family reported how one twin, placed on report (a disciplinary measure) for poor behavior, had not put effort into English for some time. A new teacher arrived in Year 10 (the first year of the GCSE course) who played rap music in order to interest students in literature and poetry. This previously disaffected twin said lessons became more interesting and that, as a result, he behaved better, worked harder, and was taken off report. His co-twin was said to have been more interested and hardworking throughout. Although the co-twin performed better in GCSE (A and B compared with two Cs), the improvement in behavior and higher-than-expected grades for the previously disaffected twin were attributed by this family to his new English teacher.

Issues around teachers being unable or unwilling to control their class featured throughout the interviews, not only in the context of ability grouping. One twin was said to have had a poor math teacher in Year 9 who struggled with behavior management and at the end of this year was predicted a Grade E. However, partway through Year 10, the class got a new teacher, and at the end of Year 11, this twin achieved a B, three grades higher than predicted.

> [The new math teacher] had a very stern approach to things. A firm hand. He was strict but fair. You would always do your homework; you would always work hard in the class. . . . [The previous teacher] was notoriously bad. . . . He didn't have any control, didn't have any control over anybody . . . and everybody used to talk and he didn't seem to mind everybody talking which was bizarre because he was a teacher and should have been teaching us.

It is noteworthy, however, that this pupil's co-twin achieved an A*. As with other families, the achievement difference was explained by perceived differences in teacher quality.

Perceived teacher–pupil relationship. Participants talked about the importance of the teacher–pupil relationship but sometimes in contradictory ways. In cases where the relationship was poor, some students reported feeling demotivated and giving up on the subject and therefore performing less well than their co-twin. Other students, however, felt that a poor relationship with the teacher pushed them to work harder to prove their teacher wrong:

> Well, I know that I didn't really get on with my teacher much, and she told me things like I would be lucky if I got a C. . . . We just didn't gel. She was kind of a bit like that with everyone really, apart from the ones she really liked. . . . I worked quite hard . . . I wanted to prove her wrong. That was the only motivation I had.

This pupil succeeded and achieved a grade B. However, it must still be noted that her co-twin, who did not experience a problematic teacher–pupil relationship, achieved an A*. The twins put this difference down to their different experiences with their English teachers.

Some families referred to the impact of teachers having "favorite" students. For instance, in one family, despite both twins being in the same mathematics set, their experience was seen as markedly different. One of the twins reported,

> I hated him . . . the teacher must definitely be part of it. You would stick your hand up and he wouldn't even come to you. He would just choose favorites, it was ridiculous. The person [co-twin] sat next to was one of his favorites so he was always on that table, which obviously helped.

The chance event of one twin sitting next to one of the teacher's "favorite" students and thus indirectly receiving more support was seen as the main cause of the twin quoted above achieving a grade D in math while his co-twin achieved a B.

In both interviews and questionnaires, participants not only attributed achievement discordance to the school environment—and in smaller numbers to factors such as bullying, being distracted by social media, and romantic relationships—they also explained discordance with reference to traits or behavior.

Nonshared Effects of Discordant Traits or Behavior. Many families, when asked for their explanations of discordant GCSE results, mentioned differences in traits and behavior rather than specific environments. These cannot be considered as NSE influences themselves but must be the result of nonshared experiences. Follow-up interviews made it possible to ask families to explain when and why one twin began to diverge from the other in terms of effort, interest, ability, and personality.

Effort. Effort was the most commonly cited explanation for discordant GCSE results. Parents and twins alike reported that the twin who worked harder or revised more for assessments performed better. Interviewers probed for explanations for within-pair discordance in effort, and families spoke of the influence of the twin relationship, peer relationships, and plans for the future.

One participant argued that he was more competitive than his twin and was driven by wanting to outperform him:

> I think it was possibly a competition thing. I believe I compete more with [twin] than [twin] competes with me sometimes. . . . There's always been, obviously, friendly competition between brothers. It was only at GCSE that we were split up completely in terms of academically—and I think that's possibly where it all came from. When we split up, I thought I'll try and be my own . . . and push a bit further. It does sound horrible—leave him in the dust a bit.

However, as the interview progressed, it transpired that this twin had been bullied and had a history of self-harming. One of the reasons he studied harder than his twin was that he did not have good friends or a social life. He was motivated by wanting to "improve himself" and to get good enough grades to be able to leave school to go to college. In addition, the career path that appealed to him required good grades in academic subjects, something that was not the case for his co-twin. The harder-working twin performed better than his co-twin and achieved, for example, a grade A in English compared with his brother's grade C.

Issues with peers were mentioned by other families in relation to discordant motivation. For example, one twin had been in top set for math in Year 7 but was dropped down to a lower set in Year 8. This twin reported that having no friends in the new set motivated her to work harder and get herself moved back up by Year 9. She achieved an A* grade at GCSE while her twin achieved a B. In another example, one twin reported revising more for GCSE Physics partly because he liked the subject but also because friends in his boarding school house studied physics and they revised together. He achieved a grade A whereas his co-twin achieved a C. Another twin reported enjoying English classes and working hard at English throughout secondary school, whereas his co-twin did not enjoy the subject and messed around with friends during class. While he achieved a grade A, his twin achieved a grade C. It can be seen that being in a group with friends was not always perceived as having the same effect.

Some young people reported not putting as much effort into subjects they expected would not be of use to them in their careers. For example, one twin wanted to work with children and was not motivated to work at math, which she did not perceive as relevant to her goal. However, on realizing that she may not pass her math GCSE (and therefore not be accepted onto her chosen course), she sought extra help and spent extra time on math revision. Ultimately she achieved a C grade. By contrast, her twin who had studied hard at math throughout the course achieved an A grade.

Interest. Parents and twins in several families explained discordance in GCSE results on the grounds of the twins having different levels of interest. One twin who loved math was awarded an A*, whereas her twin sister, who was less positive about math, achieved a C. However, other than reports of inspirational teachers triggering interest, few environmental explanations were offered for discordance in interest between twins.

Ability. In spite of the shared DNA of MZ twins, several families believed that one twin had more "natural" ability in a subject than the other and that this explained discordance in achievement. One family described how one twin had been "behind" his co-twin ever since primary school. Both twins performed well in their GCSEs, but the twin considered "behind" got Cs in English language and literature, whereas his co-twin achieved an A and a B. In math, he achieved a B and his brother an A*. Environmental explanations were rarely offered for perceived ability differences.

Personality. Some families described characteristics such as self-confidence or perfectionism as reasons for one twin's performing better than the other. For instance, one twin was described as a perfectionist, a slow writer who found it difficult to get his ideas down on paper. His mother commented,

> He's always been like that, even when he was little and learning to write, if he did little notes on bits of paper and it wasn't . . . he'd scrumple it up and do another one. . . . It would always have to be just right.

The greatest GCSE grade difference between these twin boys was in English. Despite being in the same set with the same teacher, the "perfectionist" twin achieved a B and his co-twin an A*. Again, environmental explanations were not generally offered for personality discordance.

Discussion

This study was designed to generate testable hypotheses about nonshared experiences that may influence GCSE achievement. The data suggested two main areas from which hypotheses, to be tested in future work, might be drawn: MZ discordance in school experiences and MZ discordance in behavioral traits.

School Experiences

The differential experiences offered most regularly to explain discordant achievement were ability grouping, perceived teacher quality, and perceived teacher–pupil relationships.

In general, a twin placed in a higher-ability group performed better than his or her co-twin in a lower group. This could be explained simply as the twin who was better at the subject being placed in a higher set and achieving a higher grade as a direct result of his or her higher ability. Our design cannot tell us about the direction of effects. However, with identical twins, it is important to ask why one twin would have shown higher ability or achievement in a particular subject as the difference cannot be due to genetic differences. It is clear from both questionnaire and interview data that participants saw differential ability grouping as a cause, not just a consequence, of achievement discordance. They

described better teaching, higher teacher expectations, better behavior, and more motivated pupils as positive influences in higher sets.

These lay explanations are in line with existing research into the effects of ability grouping on attainment both in general (e.g., Slavin, 1990) and for GCSE specifically. In a large study of $n = 6{,}000$ pupils in 45 U.K. comprehensive schools, no overall effect of ability grouping on GCSE achievement was found after other factors, including gender, prior attainment, social disadvantage, and attendance, were taken into account (Ireson, Hallam, & Hurley, 2005). However, on closer inspection, it was seen that students of similar prior attainment (midlevel achievers were dispersed throughout the full range of ability groups) performed better in top than in middle sets and better in middle than in lower sets. The authors concluded that despite the lack of a general effect, it is possible that individual students were affected by the set in which they were placed.

This finding strengthens the hypothesis emerging from our qualitative data that ability grouping may act as an NSE influence on GCSE achievement. Ireson et al. (2005) argue that factors such as different curricula and pedagogical approaches, teacher attitudes and expectations, student motivation, and not being placed in the most appropriate ability group could all mediate the impact of ability grouping on achievement. Several of these factors were mentioned explicitly by our participants. Alignment between the current qualitative study and extant quantitative research strengthens the hypotheses offered by the twins and their parents and suggests that we are not just experiencing noise around a weak signal.

Findings suggest that ability grouping, and being moved between sets, could explain some NSE variance in GCSE achievement. This hypothesis will be tested empirically in the next step of the current research program. If ability grouping can explain variance in achievement, independent of genetic effects, then implications for school organization may be considered. However, it will be important to bear in mind that different pupils responded differently to discordant ability grouping. This highlights the importance of personalizing educational decisions rather than taking a one-size-fits-all approach.

Families also explained discordance in achievement by saying that one twin had a better teacher than the other. Student perceptions of teacher quality have been explored and overviews of findings presented in reviews in both the United Kingdom (Coe, Aloisi, Higgins, & Major, 2014) and the United States (Bill and Melinda Gates Foundation, 2012). Coe et al. (2014) reviewed available evidence on student ratings as a tool for gauging teacher effectiveness and offering formative feedback. They presented evidence that student ratings have been found to be both internally and externally valid and reliable, a finding reiterated by the authors of the Gates Foundation's 2012 report, who commented, "No one has a bigger stake in teaching effectiveness than students. Nor are there any better experts on how teaching is experienced by its intended beneficiaries" (Bill and Melinda Gates Foundation, 2012, p. 2). Our study allowed us to question some of these "experts" in detail about the perceived quality of their GCSE teachers.

The Gates Foundation report found that student survey results were predictive of achievement gains and yielded more consistent results than classroom observation, supporting our decision to explore perceived learning environments (Bill and Melinda Gates Foundation, 2012). It seems fair to say that the available evidence suggests that student perceptions of teacher quality can be reliable and valid predictors of student achievement. This was reflected in what a proportion of the families in the current study told us. A testable hypothesis is suggested that students' perceptions of how good their teacher is could explain NSE variance in GCSE achievement. As with ability grouping, it is important to bear in mind that in the real world, there will always be variability in teacher quality and in perceptions of how good individual teachers are. The "best" teachers will not necessarily be best for all students. We will need to dig deep to identify precisely what it is in teachers' behavior that students respond to and the extent to which responses are consistent from student to student.

The final "environmental" hypothesis emerging from the data related to perceptions of teacher–pupil relationships, an aspect of experience that has been researched extensively and found to relate to a range of outcomes over several decades. One U.S. national survey of adolescents found that teachers were commonly listed when young people were asked to identify the emotionally supportive relationships in their lives (Resnick et al., 1997). Harter (1996) described how the relationship between child and teacher changes developmentally, becoming more formal, evaluative, and competitive as children get older. Harter also pointed out that young people who are low in intrinsic motivation may be more negatively affected by this change than others, perhaps explaining some of the different reactions described in the Results section of this paper.

Pianta, Hamre, and Stuhlman (2003) provided a comprehensive review of teacher–pupil relationship research that has since been updated to reflect recent developments in conceptualization and empirical evidence in the field (Sabol & Pianta, 2012). A clear picture emerges of a bidirectional relationship, linked to individual teacher and pupil characteristics, that correlates with teacher behavior and student outcomes, including trajectories to academic success or failure (Hamre & Pianta, 2001; Wentzel, 1998). Much of the research has focused on young children and on teacher perspectives or independent observations, but there is a smaller body of evidence on older pupils' perceptions of their relationship with teachers. Attachment, developmental systems theory, social-motivational theory, interpersonal theory, socialization, and social support models have all been used to describe teacher–pupil relationships in adolescence (Pianta & Allen, 2008). Sabol and Pianta (2012) argued that perceptions of emotional support and relatedness are central to all of these models of teacher–pupil

relationships in adolescence, factors that come across clearly in responses from the adolescent twins and their parents in the current study. Recent research has also found that teacher–pupil relational skills can be taught and can lead to improved student achievement, a finding with clear implications for teacher training (Allen, Pianta, Gregory, Mikami, & Lun, 2011; Hamre et al., 2012). It would be unreasonable to expect the same relationship between a teacher and all of his or her pupils. However, the data suggest that if a pupil has a particularly strong dislike of his or her teacher, or perceives the relationship to be very poor, then this perception might explain some NSE variance in achievement in the GCSE subject taught by that particular teacher.

Behavioral Differences

Many families, when asked why one twin achieved more than the other at GCSE, offered behavioral rather than environmental explanations. In particular, they cited differential effort, ability, interest, and personality. Of these, the most commonly cited explanation was effort, and this was the only behavior for which twins and their parents offered environmental explanations. Differences in ability and personality were usually described rather than explained, and other than a mention for inspirational teachers, the same was true for interest. It would be interesting to pursue why families had so few explanations for this type of discordance. One possibility is that the differences have been there for as long as they remember and reflect pre- or perinatal NSE influences. However, this seems unlikely to be the case for interest, and the area merits further and more in-depth exploration.

A large proportion of families told us that the twin who performed better also worked harder. One explanation offered for this involved knowledge of the entry requirements for future employment and study. The idea that pupils may be motivated by knowledge of the academic entry requirements for careers that interest them seems like fertile ground for intervention-focused hypotheses that could be pursued in future research.

Peers were also mentioned in relation to discordant effort in both positive and negative ways. Being in lessons or study groups with friends was seen as a positive influence on achievement, whereas bullying and messing about in class were seen as negative influences. Peer relationships as a potential NSE influence on motivation merits further research.

Future Research

These findings suggest that some aspects of teaching and school organization may act as environmental influences on academic achievement that are, to some extent, uncorrelated with genes. The ideas generated by this qualitative study require quantitative testing. The data presented here are being used to develop a quantitative measure of NSE influences at age 16, which will be rooted in the explanations of discordance offered by these adolescent twins and their parents. It will be possible to assess whether the experiences identified can explain NSE variance in U.K. 16-year-olds' GCSE achievement. It will also be interesting to explore genotype–environment correlations using this new measure.

Future research will ask how much of the NSE variance in GCSE achievement can be explained by perceptions of teacher quality and teacher–pupil relationships and by ability grouping. We need to know to what extent within-family effects are also observed between families. The full MZ-dizygotic twin design is useful for this purpose as we can explore the extent to which associations between two variables, such as ability grouping and GCSE achievement, are mediated by NSE. If variables identified by this study are found to account for individual differences in achievement, we may be able to consider their implications for educational practice. If effects are small, as would be predicted by extant research, we will also be able to explore whether they can be combined in any meaningful way to represent accumulated environmental risk or prediction, similar to the polygenic risk predictors currently being identified in molecular genetic research (e.g., Okbay et al., 2016). It will be interesting to explore whether, for instance, how having a high level of interest and also a particularly good teacher, or a high level of neuroticism and a peer relationship problem, affects achievement.

Studies could also explore individual differences and changes in motivation after being moved up or down an ability group. Further research into environmental influences on how hard pupils work, perhaps starting with careers education and peer relationships, represents another promising line of inquiry. Finally, looking for NSE influences on the other behavioral traits mentioned by participants—ability, interests, and personality—represents an interesting direction for further research.

Our expectation is that, as is the case with genes, each experience is likely to have only a small effect. Furthermore, any effects may not be stable over time (Tucker-Drob & Briley, 2014) and may interact with each other and with genotypes. We would argue that identifying environmental experiences that explain any variance in GCSE is a useful endeavor, even if only to U.K. 16-year-olds taking GCSE. This is particularly true if carefully measured NSE influences, rooted in evidence, can be used as a basis for constructive intervention.

Limitations

Although our design is uniquely powerful for identifying potential NSE candidates, it is essentially a series of case studies, and therefore, any conclusions regarding causation cannot be drawn. Furthermore, it is likely that there are some twin-specific effects at play. It is also important to restate that estimates of NSE variance include variance attributable to measurement error. This is a major limitation of our research because if measurement error can explain ~20% of the variance in GCSE achievement, then we are barking up

the wrong tree. This is not impossible as it has been estimated that National Curriculum tests can show reliabilities of around 0.80 (Wiliam, 2001). Therefore, as discussed, the amount of variance in GCSE performance that we are endeavoring to explain will certainly be less than 20% and possibly substantially less. However, if our aim is to support young people in doing as well as they are capable of doing, then every little bit helps. Everything we are learning from both quantitative and molecular genetics suggests that just as genetic effects are many, small, and involved in dynamic interplay, the same is likely to be true for environmental influence on human behavior. Furthermore, it is likely that some of our twin pairs represent outliers and that the information they have provided is unlikely to generalize to others. Unique (nonshared) experiences may be "too unique" to have meaning beyond the individuals affected by them, as suggested by Turkheimer and Waldron's (2000) review. This is certainly a possibility and indeed is very likely to be true in the most extreme cases of discordance in our data set. However, in the current paper, we have focused on experiences mentioned by at least three (and up to 22) families. The type of discordance we have focused on (two or more GCSE grades) is also relatively modest. In this way, we have attempted to focus on explanations or hypotheses with the greatest chance of being general rather than unique NSE explanations. Quantitative testing of these hypotheses will be required to assess whether they do indeed explain some of the NSE variance in academic achievement at age 16.

Conclusions

A qualitative, hypothesis-generating MZ-twin differences approach to pinning down NSE has merit in helping us identify where to look for potential NSE influences. It may be that this research strategy could be adopted more widely by behavioral geneticists as it is clear that we need to analyze human experience in as much detail as the human genome. The current study identified hypotheses about ability grouping, perceived teacher quality, and perceived teacher–pupil relationships as potential NSE influences on academic achievement in adolescence that can be tested in future work. Although previous research has found links between environmental factors emerging from this data set and academic achievement, the current study is novel in identifying them as potential candidate NSE influences.

Acknowledgments

This research was supported by a grant (EDU/40881) from the Nuffield Foundation to the first and last authors, and we are hugely grateful for their support. The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at www.nuffieldfoundation.org. We would also like to thank Twins' Early Development Study (TEDS) families for their generous participation and Andy McMillan and Rachel Ogden for their help and support in collecting and managing the data for both of these studies. Finally, particular thanks is owed to Mrs. Patricia Busfield for her expert interviewing of TEDS families.

References

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, *333*(6045), 1034–1037.

Asbury, K., Almeida, D., Hibel, J., Harlaar, N., & Plomin, R. (2008). Clones in the classroom: A daily diary study of the nonshared environmental relationship between monozygotic twin differences in school experience and achievement. *Twin Research and Human Genetics*, *11*(6), 586–595.

Asbury, K., Dunn, J., Pike, A., & Plomin, R. (2003). Nonshared environmental influences on individual differences in early behavioral development: A monozygotic twin differences study. *Child Development*, *74*(3), 933–943.

Asbury, K., Dunn, J., & Plomin, R. (2006a). Birthweight-discordance and differences in early parenting relate to monozygotic twin differences in behaviour problems and academic achievement at age 7. *Developmental Science*, *9*(2), F22–F31.

Asbury, K., Dunn, J., & Plomin, R. (2006b). The use of discordant MZ twins to generate hypotheses regarding non-shared environmental influence on anxiety in middle childhood. *Social Development*, *15*(3), 564–570.

Asbury, K., & Plomin, R. (2013). *G is for genes: The impact of genetics on education and achievement*. New York, NY: Wiley.

Bill and Melinda Gates Foundation. (2012). *Asking students about teaching: Student perception surveys and their implementation*. Seattle, WA: Author. Retrieved from http://www.metproject.org/downloads/Asking_Students_Practitioner_Brief.pdf

Burt, S. A., McGue, M., Iacono, W. G., & Krueger, R. F. (2006). Differential parent–child relationships and adolescent externalizing symptoms: Cross-lagged analyses within a monozygotic twin differences design. *Developmental Psychology*, *42*(6), 1289–1298.

Calvin, C. M., Deary, I. J., Webbink, D., Smith, P., Fernandes, C., Lee, S. H., . . . Visscher, P. M. (2012). Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. *Behavior Genetics*, *42*(5), 699–710.

Caspi, A., Moffitt, T. E., Morgan, J., Rutter, M., Taylor, A., Arseneault, L., . . . Polo-Tomas, M. (2004). Maternal expressed emotion predicts children's antisocial behavior problems: Using monozygotic-twin differences to identify environmental effects on behavioral development. *Developmental Psychology*, *40*(2), 149.

Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). *What makes great teaching? Review of the underpinning research*. London, UK: Sutton Trust.

Csikszentmihalyi, M., & Csikszentmihalyi, I. S. (Eds.). (1988). *Optimal experience: Psychological studies of flow in consciousness*. Cambridge, UK: Cambridge University Press.

Deater-Deckard, K., Pike, A., Petrill, S. A., Cutting, A. L., Hughes, C., & O'Connor, T. G. (2001). Nonshared environmental processes in social-emotional development: An observational

study of identical twin differences in the preschool period. *Developmental Science*, *4*(2), F1–F6.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, *72*(2), 625–638.

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., . . . Scott-Little, C. (2012). A course on effective teacher-child interactions effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, *49*(1), 88–123.

Harter, S. (1996). Teacher and classmate influences on scholastic motivation, self esteem and level of voice in adolescents. In J. Juvonen & K. R. Wentzel (Eds.), *Social motivation: Understanding children's school adjustment* (pp.11–42). Cambridge, UK: Cambridge University Press.

Haworth, C., Davis, O. S., & Plomin, R. (2013). Twins Early Development Study (TEDS): A genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Research and Human Genetics*, *16*(01), 117–125.

Iervolino, A. C., Pike, A., Manke, B., Reiss, D., Hetherington, E. M., & Plomin, R. (2002). Genetic and environmental influences in adolescent peer socialization: Evidence from two genetically sensitive designs. *Child development*, *73*(1), 162–174.

Ireson, J., Hallam, S., & Hurley, C. (2005). What are the effects of ability grouping on GCSE attainment? *British Educational Research Journal*, *31*(4), 443–458.

Kendler, K. S., & Halberstadt, L. J. (2013). The road not taken: life experiences in monozygotic twin pairs discordant for major depression. *Molecular Psychiatry*, *18*(9), 975–984.

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J. B., . . . Plomin, R. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences*, *111*(42): 15273–15278.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Oskarsson, S. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*, 539–542.

Oliver, B. R., Pike, A., & Plomin, R. (2008). Nonshared environmental influences on teacher-reported behaviour problems: Monozygotic twin differences in perceptions of the classroom. *Journal of Child Psychology and Psychiatry*, *49*(6), 646–653.

Oliver, B. R., & Plomin, R. (2007). Twins' Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. *Twin Research and Human Genetics*, *10*(01), 96–105.

Pianta, R. C., & Allen, J. P. (2008). Building capacity for positive youth development in secondary school classrooms: Changing teachers' interactions with students. *Toward Positive Youth Development: Transforming Schools and Community Programs*, *1*, 21–40.

Pianta, R. C., Hamre, B., & Stuhlman, M. (2003). Relationships between teachers and children. *Handbook of Psychology*, 10(3),199–234.

Pike, A., Reiss, D., Hetherington, E. M., & Plomin, R. (1996). Using MZ differences in the search for nonshared environmental effects. *Journal of Child Psychology and Psychiatry*, *37*(6), 695–704.

Plomin, R. (2011). Commentary: Why are children in the same family so different? Non-shared environment three decades later. *International Journal of Epidemiology*, *40*(3), 582–592.

Plomin, R., & Daniels, D. (1987). Why are children in the same family so different from one another? *Behavioral and Brain Sciences*, *10*(01), 1–16.

Price, T. S., Freeman, B., Craig, I., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research*, *3*(03), 129–133.

Reiss, D., Hetherington, E. M., Plomin, R., Howe, G. W., Simmens, S. J., Henderson, S. H., . . . Law, T. (1995). Genetic questions for environmental studies: Differential parenting and psychopathology in adolescence. *Archives of General Psychiatry*, *52*(11), 925–936.

Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., . . . Udry, J. R. (1997). Protecting adolescents from harm: Findings from the National Longitudinal Study on Adolescent Health. *JAMA*, *278*(10), 823–832.

Rimfeld, K., Kovas, Y., Dale, P. S., & Plomin, R. (2015). Pleiotropy across academic subjects at the end of compulsory education. *Scientific Reports*, *5*, 11713.

Ritchie, J., & Spencer, L. (1994). Qualitative data analysis for applied policy research. In A. Bryman & R. Burgess (Eds.), *Analysing qualitative data* (pp. 173–194). London, UK: Routledge.

Sabol, T. J., & Pianta, R. C. (2012). Recent trends in research on teacher–child relationships. *Attachment & Human Development*, *14*(3), 213–231.

Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Rimfeld, K., Krapohl, E., Haworth, C. M. A., & Plomin, R. (2013). Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PLOS ONE*, *8*(12), e80341.

Slavin, R. E. (1990). Ability grouping in secondary schools: A response to Hallinan. *Review of Educational Research*, *60*(3), 505-507.

Tucker-Drob, E. M., & Briley, D. A. (2014). Continuity of genetic and environmental influences on cognition across the life span: A meta-analysis of longitudinal twin and adoption studies. *Psychological Bulletin*, *140*(4), 949.

Turkheimer, E., & Waldron, M. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin*, *126*(1), 78.

Viding, E., Fontaine, N. M., Oliver, B. R., & Plomin, R. (2009). Negative parental discipline, conduct problems and callous–unemotional traits: Monozygotic twin differences study. *British Journal of Psychiatry*, *195*(5), 414–419.

Vitaro, F., Boivin, M., Brendgen, M., Girard, A., & Dionne, G. (2012). Social experiences in kindergarten and academic achievement in Grade 1: A monozygotic twin difference study. *Journal of Educational Psychology*, *104*(2), 366.

Wentzel, K. R. (1998). Social relationships and motivation in middle school: The role of parents, teachers, and peers. *Journal of Educational Psychology*, *90*(2), 202.

Wiliam, D. (2001). Reliability, validity, and all that jazz. *Education 3-13*, *29*(3), 17–21.

Authors

KATHRYN ASBURY is a twin researcher who lectures in Psychology in Education at the University of York, UK.

NICOLA MORAN is a research fellow in the Department of Social Policy and Social Work at the University of York, UK.

ROBERT PLOMIN is professor of behavioral genetics in the Social Genetic and Developmental Psychiatry Centre, King's College London, UK.