Exploring when maternal interest is sufficient for high attainment in mathematics: A configurational analysis using longitudinal data

Stephanie Thomson

School of Education, Durham University

Abstract

Qualitative Comparative Analysis (QCA) is a case-based method, developed by Ragin (1987, 2000), to analyse medium- and large-n datasets. It uses Boolean algebra to show which configurations of factors in a model are either necessary and/or sufficient for a specified outcome. In the social world, we rarely see perfect necessity and sufficiency but we can use QCA to assess the degree of necessity or sufficiency to find configurations which are quasi-necessary or quasi-sufficient. In this paper, I use crisp-set QCA on data from the 1970 Birth Cohort Study (BCS70) to investigate which configurations of sex, maternal interest, social class and, later, ability are quasi-sufficient for various levels of attainment in maths. Firstly, I explain how to conduct QCA, through the use of examples, before using a set-theoretic measure of consistency to explore the relationship between sex, social class, maternal interest and, what I term, above-average attainment in mathematics. To this model, I then introduce an additional factor of general ability (operationalised as several dichotomous factors, each indicating a certain level of ability) leading to instances of configurations having strong subset relations but containing very few cases. These rows, called remainders, cannot be included in a solution without theoretical justification (Ragin, 2008). For the final stage of the analysis, I create, for two different general ability levels, a ‘most-complex solution’ (which excludes all remainder rows) and a parsimonious solution (which includes any remainder row contributing to parsimony). These act as boundaries for the ‘intermediate solution’ which contains only those remainders which can, theoretically, be thought to obtain the outcome. I then discuss each intermediate solution and note that, in one case, it is the same as the relevant most-complex version.

Keywords: Education, Qualitative Comparative Analysis (QCA), social class, mathematics, parents, parental interest

Introduction

This paper uses data from the 1980 sweep of the 1970 Birth Cohort Study (BCS70) to examine the mathematics attainment of 10-year-old children. I use the case-based, set theoretic method Qualitative Comparative Analysis (QCA) to assess to what extent maternal interest in education is sufficient for high attainment in mathematics. More specifically, I examine whether maternal interest is sufficient only for children of certain social classes and a particular sex. I choose to use QCA because it assumes that ‘multiple conjunctural causation’ is at work rather than an average ‘net-effect’ of each variable across all cases (Cooper and Glaesser, 2008). This allows us to explore which combinations of factors – or configurations –
consistently achieve the outcome and, hence, account for factors whose effects are a ‘function’ of other factors in the configuration (Cooper and Glaesser, 2008).

Here, following the approach of Cooper (2005) and Cooper and Glaesser (2009), I will use fs/QCA software (developed by Ragin) to conduct QCA on a large dataset to show which configurations (of the sex of the child, parental interest and the social class of parent(s)) typically lead to high attainment in mathematics. I will then refine the model through the introduction of another factor – the ability of the child - in an attempt to reduce the number of “contradictory” configurations where high or low mathematics attainment is equally likely.

This paper is part of a larger, doctoral study investigating how parental involvement in (and its effects on) mathematics differs by social class. Desforges (2003) suggests that parental involvement, conceived of as ‘at-home interest and support’, impacts positively on attainment regardless of the sex or social class of the child. The QCA results reported here focus on sex and class related differences in the effect of involvement and are being used to inform case-selection for interviewing which aims to unpack differences in the type of involvement. In BCS70, I am not able to distinguish between types of involvement directly but I can explore, for example, whether, for some children of a particular social class, having an interested mother is sufficient for high achievement, irrespective of a child’s level of general ability.

The model incorporating ability contains some configurations with very few cases and I discuss whether these ‘remainder’ rows can be incorporated safely into any simplified version of a solution. Through the use of two examples, I show how counterfactual reasoning can be used to include (or exclude) individual rows with a low number of cases.

**Literature**

A central assumption of methods dealing with conjunctural causation is that causes do not act independently. Much of the previous large-n work on parental involvement in education and attainment inequalities proceeds using regression-based methods. Research in this vein will typically involve looking for correlations between high levels of interest from parents and achievement by children or searching for common attributes of different parents who have high levels of interest in their children’s schooling (see, for example, Bakker, Denessen, & Brus-Laeven, 2007; Domina, 2005; Friedel, Cortina, Turner, & Midgley, 2007; Hirata, Nishimura, Urasaka, & Yagi, 2006).

Standard statistical methods which look for the ‘unique contribution’ (Desforges, 2003) of parental involvement to, say, attainment assume, in the first instance, that a unique contribution can be found and that searching for it is helpful to analysis. Statistical methods use, what Ragin (2000) terms, ‘net-effects thinking’ and focus on the effect of one variable independent of the values of other variables. This also means they may fail to draw a qualitative distinction between cases that differ, for example, on only one variable in a model. To pick apart our cases into their constituent variables would be an unhelpful way to begin to explain any complex causation.

Much existing work investigating parental involvement using broadly case-based methods consists of analysis of interview data. Reay (1998) interviewed a sample of mothers to explore how their social class background influenced the type of help they were able to give with homework and Crozier (1999) interviewed parents to ask about their relationship with the school. Both these pieces of work, typical of those in this vein, provide a detailed account of the differences in the type of involvement engaged in by different parents but steer away from a discussion of attainment.

Attainment, however, is of key importance in mathematics. Traditionally, mathematics has been used as a filter subject in the UK. Precisely, mathematical questions were a large part of the entrance exams for grammar schools and many university courses (and then graduate employers) ask for a good GCSE pass in mathematics as a requirement.
Complex causation

In this paper, I re-frame the discussion of parental involvement to focus on cases and the potential causal factors within a case. Whereas many researchers working with variables and regression techniques would expect their variables to act in a uniform way across all cases, case-based researchers expect their qualitative factors to interact with other factors in the case (e.g., Cooper & Glaesser, 2008).

This shift in thinking allows us to investigate which factors (or configurations of factors) are necessary for the outcome and which are sufficient. A necessary condition for an outcome must, as the name suggests, be present for the outcome to be achieved but may not be solely sufficient for the outcome to be achieved (Ragin, 2000). Similarly, if every instance of a factor is followed by the outcome, we can say that that factor is a sufficient condition though, again, it may not be necessary. In some cases, a factor (or configuration) may be a necessary and sufficient condition and, in others, it may be neither necessary nor sufficient on its own.

The set of cases with a sufficient condition, when conceived in set theoretic terms, is a subset of the set of cases with the outcome. Similarly, the set of cases with the outcome is a subset of the set of cases with a necessary condition. When examining the social world, however, it is unusual to encounter such examples of perfect necessity or sufficiency. The concepts themselves are still a useful way to unpack complex causation, however, because they allow us to determine whether a factor is, for example, ‘almost always’ sufficient (Ragin, 2000). Such a factor could be deemed a quasi-sufficient condition. Of course, terminology like ‘quasi-sufficiency’ on its own is not enough to describe the degree of sufficiency and we need both a way of calculating the degree of sufficiency and deciding what degree counts as quasi-sufficient. These terms (and how to calculate them) are discussed below. In this paper, all the analyses address sufficiency.

How to conduct Qualitative Comparative Analysis

Ragin (1987) originally developed crisp set Qualitative Comparative Analysis (csQCA) as an alternative to variable-based methods for small- to medium-n datasets. The mathematical principles underlying it, however, mean that it also provides a rigorous way to search for (quasi-)sufficient factors (or configurations of factors) in large datasets. With the hypothetically causal conditions and an outcome measure specified, software called fs/QCA generates truth tables with a row representing each possible configuration of factors. Contained in that row is information about the number of cases in each configuration and a ‘consistency’ score.

This consistency score, in csQCA, is a value between 0 and 1 which shows the proportion of cases in the configuration with the outcome. The higher the consistency score, the higher the proportion of cases with that configuration that have the outcome. High consistency scores should, however, be viewed in conjunction with the number of cases as low numbers can sometimes produce consistency scores which are misleading. Researchers set a threshold for consistency during analysis and include the rows above the threshold in the solution. The fs/QCA software then uses Boolean algebra to produce a simplified version of this solution of the form: \( Y = (A \times b \times c) + (C \times D \times E) \), where \( Y \) is the outcome measure and \( A, B, C, D \), and \( E \) are factors.

This notation allows the reader to see quickly (once familiar with the concepts) what may be happening in the dataset. This solution has 2 terms, contained in brackets, which can each be thought of as routes to the outcome (Ragin, 2008). The notation uses upper case to show a factor is present and lower case to show it is absent. The symbol * is used to indicate the intersection of the sets in question (often referred to as “logical AND”) and the symbol + is used to indicate set union (also called “logical OR”). So, in this example, one route to the outcome is for the configuration to contain factors \( A \) and not factors \( b \) or \( c \). Notice that this term makes no stipulations about factors \( D \) or \( E \). Another route to the outcome is for the configuration to contain factors \( C, D \), and \( E \). This solution does not tell us that these are the only two possible routes to the outcome but that, given a well-chosen consistency threshold, a high proportion of cases described by either term will obtain the outcome.
We want the consistency threshold to be high enough (usually 0.75 and above) to ensure almost all of the cases in a row above the threshold obtain the outcome but it is not enough to pick a value and blindly exclude all rows below it. We must, at the same time, group together rows with negligible differences in consistency. Table 1 shows an example of a truth table whose rows have been ordered by consistency. We can see that rows 3, 4 and 5 have negligible differences in consistency and it would be misleading to include only one or two of these rows in the solution. So, here, we should choose to set the consistency threshold at 0.8 and exclude all 3 or 0.72 (and include all 3). This final judgement depends on what level of sufficiency is appropriate in a particular piece of research.

**Table 1: Example of a Truth Table**

| A | B | C | number | consistency |
|---|---|---|--------|-------------|
| 1 | 1 | 1 | 450    | 0.89        |
| 0 | 0 | 1 | 691    | 0.82        |
| 1 | 1 | 0 | 370    | 0.76        |
| 1 | 0 | 0 | 114    | 0.73        |
| 0 | 1 | 1 | 16     | 0.72        |
| 1 | 0 | 1 | 349    | 0.58        |
| 0 | 0 | 0 | 208    | 0.47        |
| 0 | 1 | 0 | 84     | 0.12        |

If we think of each term in a solution as being one route to the outcome, it is helpful to know the ‘empirical importance’ of each of these routes (Ragin, 2008). We find this, in the crisp-set context, by calculating the ‘coverage’ of a configuration by dividing the number of cases in it which obtain the outcome by the total number of cases in the sample which obtain the outcome.

When interpreting a piece of fs/QCA output, we are given two (potentially different) values for coverage. For each term, we have a ‘raw coverage’ score which tells us the gross coverage of the term and does not take into account whether any configurations included in the term also occur in another term. The ‘unique coverage’ of a term allows us to see the proportion of the outcome which is covered by the parts of that term which do not overlap with other terms.

In the illustrative scenario in Table 2, we see all the possible configurations of the factors ‘A’, ‘B’ and ‘C’ and ‘D’, their consistencies and how many cases are represented by each. If we include all the rows marked ‘quasi-sufficient’ in the solution, we get the output in Figure 1. Looking at Figure 1, we see that ‘b*D’ accounts for approximately 28% of the outcome while ‘a*D’ and ‘C*D’ account for approximately 14% and 35% respectively. All of these terms, however, have a unique coverage of under 0.1 which means that none of them, uniquely, account for more than 10% of the outcome.
Table 2: Example truth table

| A | B | C | D | number of cases | consistency | quasi-sufficient? |
|---|---|---|---|-----------------|-------------|------------------|
| 0 | 1 | 1 | 1 | 44              | 0.98        | yes              |
| 1 | 0 | 1 | 1 | 184             | 0.95        | yes              |
| 0 | 0 | 1 | 1 | 110             | 0.95        | yes              |
| 0 | 0 | 0 | 1 | 13              | 0.92        | yes              |
| 1 | 1 | 1 | 1 | 87              | 0.92        | yes              |
| 0 | 1 | 0 | 1 | 10              | 0.9         | yes              |
| 1 | 0 | 0 | 1 | 28              | 0.89        | yes              |
| 0 | 0 | 1 | 0 | 159             | 0.74        | no               |
| 1 | 0 | 1 | 0 | 164             | 0.69        | no               |
| 0 | 1 | 1 | 0 | 161             | 0.67        | no               |
| 0 | 0 | 0 | 0 | 53              | 0.6         | no               |
| 1 | 0 | 0 | 0 | 83              | 0.6         | no               |
| 1 | 1 | 0 | 1 | 15              | 0.6         | no               |
| 1 | 1 | 1 | 0 | 191             | 0.54        | no               |
| 0 | 1 | 0 | 0 | 152             | 0.47        | no               |
| 1 | 1 | 0 | 0 | 199             | 0.43        | no               |

To understand why the three terms have such low unique coverage, we must think of them as collections of rows from Table 2. The only quasi-sufficient row unique to ‘b*D’ is the row ‘A*b*c*D’ with 28 cases. Similarly, the only quasi-sufficient row unique to ‘a*D’ is ‘a*B*c*D’ with 10 cases and to ‘C*D’ is ‘A*B*C*D’ with 87 cases. We often obtain output like that in Figure 1, where some of the terms have a factor or several factors in common. Examining the raw and unique coverage scores for each term shows us how much of the outcome in general is being explained by the term and how much is uniquely covered by it.

Figure 1: Solution for Table 2, consistency threshold = 0.80

--- TRUTH TABLE SOLUTION ---

| raw coverage | unique coverage | consistency |
|--------------|-----------------|-------------|
| b*D+         | 0.278169        | 0.022007    | 0.943284    |
| a*D+         | 0.147887        | 0.007923    | 0.949153    |
| C*D          | 0.353873        | 0.070423    | 0.945882    |

solution coverage: 0.394366
solution consistency: 0.941176

The large-n dataset being used for the QCA analysis in this paper is the 1970 Birth Cohort Study (BCS70). BCS70 is a longitudinal dataset which follows a sample of people born in a particular week in 1970. Its initial questions centered on infant health but, as the participants grew older, the questions increased in scope to collect information about education, jobs and social conditions. The questions I make use of here mostly come from the 1980 sweep of BCS70 which had a particular focus on educational attainment.
The factors I will consider are mother’s interest in the child’s education, social class of the parents, sex of the child and general ability of the child against an outcome measure of mathematics attainment. I use 1662 cases in the analysis which is around 10% of the original BCS70 sample. Cases were filtered out if they had missing data in any of the variables being used to construct the factors.

The analysis was undertaken in two stages. Initially, I investigated configurations of sex, maternal interest and social class and hoped to find quasi-sufficient configurations for very-high and high attainment in mathematics. This proved unsuccessful and so I relaxed the mathematics attainment measure to indicate above-average attainment, using the same factors. At this point, I introduced a factor representing general ability into the model. This new factor doubled the number of rows in each truth table and led to some rows having very few cases. I discuss ways of treating such rows towards the end of the paper.

Factors in the model

Mathematics attainment

Mathematics attainment in BCS70 was measured using ‘The Friendly Mathematics Test’. This test was administered to the 10 year-old children and it contained 72 questions, not in order of difficulty. The test was specifically designed so that most children would be able to obtain a score above 0 and a child’s score was calculated by summing the number of correct answers. The mean score on the test in my sample was 51.39 which is higher than the mean for the whole sample, 49.35. As a result, I will have, proportionally, more cases achieving the outcome than in BCS70, as a whole. The Friendly Mathematics Test (FMT) scores were used to create various crisp outcome measures. One of these indicates whether a child is in the top scoring 50% (of the overall BCS70 cohort) for mathematics attainment or not.

I use a crisp measure, here, because I am interested in the configurations of factors which are quasi-sufficient for achieving at least a particular level of attainment in mathematics. Of course, it would be possible to create a fuzzy score for the mathematics attainment outcome but I examine the relative scores of children in the sample to attempt to understand the types of children who may attain above the median for the sample and, hence, may have greater future educational (and work-related) opportunities.

Social class

The social class categories are derived from the Registrar General’s (RG) class categories in the BCS. Those 5 class categories have been reduced to 3 by grouping together those which are most qualitatively similar. Figure 2 shows the original RG categories, their descriptions and the new categories. Particularly of note is that Class III remains split with the non-manual part constituting its own class entirely in the new scheme. As Cooper and Glaesser (2008) note, there are problems with using the RG class scheme for sociological analysis but it was the class scheme in use when the data were collected and so I use it here.

Figure 2: Collapsed class categories (from Cooper and Glaesser (2008))

| Category labels | Summary description |Collapsed category labels |
|----------------|---------------------|-------------------------|
| I              | Professional        |Professional, managerial and technical (PMT) class |
| II             | Managerial-technical|Intermediate class        |
| III1M          | Routine non manual  |Intermediate class        |
| III1M          | Skilled manual      |Intermediate class        |
| IV             | Partly skilled      |Working class             |
| V              | Unskilled           |Working class             |
To account for the 3 different crisp social classes, two social class factors were entered into the model giving the full range of configurations of each class factor with the other factors. One of these represented the PMT class, the other the working class. The 2 class factors serve as dummy variables and so, where there is a ‘0’ in both the ‘working class’ column and the ‘PMT class’ column, the case is in the intermediate class. This method produces some rows which do not, in fact, represent any real cases as there cannot, logically, be a ‘1’ in both columns. These redundant rows which have, of course, no cases are excluded subsequently from the analysis.

Maternal interest

The factor describing maternal interest comes from question J097 in the 1980 sweep of the Birth Cohort Study (BCS). It was answered by the child’s teacher and so gives his or her perspective on the involvement level of the parent. The exact question was:

“With regard to the child’s education, how concerned or interested does the mother appear to be:

- Very interested
- Moderately interested
- Very little interested
- Uninterested
- Cannot say/no parent figures”

To create the crisp measure of maternal interest, ‘Very interested’ was coded as 1 and ‘Moderately’, ‘Very little’ and ‘Uninterested’ were coded as 0. ‘Cannot say/no parent figures’ was categorised as missing data. Thus, I can qualitatively talk of a high level of maternal interest (as recorded by the child’s teacher) being represented by the presence of this factor.

Sex of the child

This information comes from the first sweep of the BCS. I call my factor ‘MALE’ here to be consistent with other pieces of literature using QCA. Hence, a ‘1’ in the male column represents a boy and a ‘0’ represents a girl.

General ability

The additional factor introduced into analyses later in the paper is an indicator of ability at age 10. This is derived from the British Ability Scale (BAS) test scores in the BCS. The mean test score for our sample was 40.75 compared with 37.58 in BCS70, as a whole. As with the FMT scores, I have created a crisp set by considering the top 50% of general ability in the BCS70 cohort with more children, proportionally, achieving these outcome levels in my sample than in BCS70 as a whole.

Initial analysis

The results below are for the initial model which does not include a factor of general ability. These are included primarily to illustrate certain methodological points of interest which arise when conducting QCA.

Above-average mathematics attainment

To consider which configurations are quasi-sufficient for above-average attainment in mathematics, I create a truth table with above-average attainment in mathematics as an outcome measure and order it by consistency. The column labelled ‘quasi-sufficient?’ is not generated by the fs/QCA software but is used to show the reader
which rows I consider quasi-sufficient for the outcome and, hence, which rows will be a part of any solution generated\textsuperscript{13}. The ‘yes’ entries in the 7\textsuperscript{th} column show that 2 rows are quasi-sufficient for this outcome.

**Table 3: Truth Table with outcome of top 50\% of mathematics achievement**

| male | working-class | PMT-class | maternal interest | number | consistency | quasi-sufficient? |
|------|---------------|-----------|-------------------|--------|-------------|------------------|
| 1    | 0             | 0         | 1                 | 96     | 0.85        | yes              |
| 1    | 0             | 1         | 1                 | 344    | 0.8         | yes              |
| 0    | 0             | 1         | 1                 | 270    | 0.71        | no               |
| 0    | 0             | 0         | 1                 | 72     | 0.6         | no               |
| 1    | 0             | 1         | 0                 | 106    | 0.58        | no               |
| 1    | 1             | 0         | 1                 | 285    | 0.56        | no               |
| 1    | 0             | 0         | 0                 | 31     | 0.52        | no               |
| 0    | 1             | 0         | 1                 | 210    | 0.51        | no               |
| 0    | 1             | 1         | 0                 | 62     | 0.42        | no               |
| 0    | 1             | 0         | 0                 | 19     | 0.37        | no               |
| 1    | 0             | 0         | 0                 | 222    | 0.27        | no               |
| 0    | 0             | 0         | 0                 | 173    | 0.21        | no               |

The top two rows have consistencies over 0.75, our suggested lower limit, and hence can be combined to produce the simplified solution in Figure 3. From it, we see that only boys with interested mothers who are not in the working-class have a quasi-sufficient route to the outcome. Looking again at Table 3, we note that this solution only incorporates 2 quasi-sufficient rows.

**Figure 3: Solution for Table 3, consistency threshold = 0.80**

--- TRUTH TABLE SOLUTION ---

| MALE*workingclass*MATERNAL INTEREST | raw coverage | unique coverage | consistency |
|-------------------------------------|--------------|-----------------|-------------|
|                                     | 0.334892     | 0.334892        | 0.813636    |

solution coverage: 0.334892
solution consistency: 0.813636

We can, then, examine the table for the negated outcome (i.e. look at the bottom 50\% of mathematics attainment, or below-average mathematics attainment) to see if it is possible to generate a simplified solution, still adhering to the consistency threshold, which incorporates more configurations.

In Table 4, there is only 1 row which could be considered quasi-sufficient – the row of working-class girls without interested mothers. We can see directly from the table, without the need for fs/QCA output, that this row has a consistency of 0.79.
In both Tables 3 and 4, there are many rows (9, in total) where the consistency figures are nearer 0.5 than either 1 or 0. This means that the configurations represented by those rows are almost equally as likely to attain the outcome as not and, hence, we cannot, from a quasi-sufficiency perspective, say much about them.

We want to be able to refine our model, by introducing an additional factor, so that a higher proportion of our rows have consistencies nearer either 0 or 1. I assumed that introducing a factor of general ability would create a new truth table with fewer rows around 0.5 and, hence, more rows which could be described as either quasi-sufficient for the obtaining the outcome or for not obtaining the outcome. Adding another factor does, however, give us truth tables with twice as many rows as those above and this, in reducing the number of cases per row, can cause added analytical problems.

**Revised model and analysis**

**Limited Diversity and Counterfactuals**

One additional problem of spreading the same number of cases over double the number of rows is an increased likelihood of having rows with very low numbers of cases in them. What we face, here, is the problem of limited diversity in the data. Ragin (2008) argues that limited diversity is a common rather than an exceptional problem when investigating the social world, even when sample sizes are reasonably large, and one which often complicates the analysis of social data.

When using QCA, we may have configurations in a truth table which are theoretically possible but empirically unlikely. We call such configurations ‘remainders’. We could, for example, create a truth table showing all the possible configurations of sex, social class, maternal interest and very-high ability (top 5%) with an outcome of very-low mathematics attainment (bottom 5%). We would expect to see very few cases overall which achieved an outcome of, say, being in the top 5% for mathematics but were also in the bottom 5% of general ability. These cases might also be spread over several configurations thus diluting their numbers further. If rows such as these contain any cases at all and at least some of these, due perhaps to sampling error, obtain the outcome, we might have rows with very high consistencies and we must, then, decide whether or not to include them in any solution14. In essence, we need to evaluate the trustworthiness of such a row’s consistency score, partly by making a theoretically-informed judgement about the likely outcome of its configuration of factors. In small- and medium-n QCA, remainders are rows with no cases at all. In the large-n context, I
consider a remainder to be a row with either no or very few cases. Rows with under 20 cases will be deemed as remainder rows as they account for (approximately) under 1% of the total cases in the sample. Other researchers working with large-n datasets may choose a different frequency threshold for remainders.

I must, first, note that including these rows in the solution may give a less complicated solution. Ragin (2008) notes that we could, somewhat blindly, choose to exclude any remainder row but that we may find our solution to be needlessly complicated as a result. Similarly, he suggests that we could include all such rows with no theoretical evaluation of their likely outcome to produce the most parsimonious solution (Ragin, 2008). Neither of these solutions is automatically preferable but they can allow us to find an upper and lower bound for the complexity of the solution.

We hope that, between these boundaries, there lies a solution including only those remainder rows which, after theoretical inspection, we think ought to be there. Such a solution is called an ‘intermediate solution’ because it sits, on a continuum of complexity, between the most complex and the parsimonious solution (Ragin, 2008). The types of remainder rows available to use in any dataset will determine how possible it is to find such an intermediate solution.

Broadly speaking, there are two types of counterfactual – ‘easy’ and ‘difficult’ (Ragin, 2008). Consider the factors A, B and C which are all thought to contribute to an outcome, X. Given the term, A*B*C, in our solution, we may think it is the presence of ‘A’ and ‘B’ alone that is producing the outcome and that ‘C’ is superfluous. We would want, then, to remove ‘C’ to simplify the solution and give a clearer summary of what is happening in the data. To do this, we would need the row A*B*C to also obtain the outcome X since, if both A*B*c and A*B*C are sufficient, we can replace these by simply A*B. Here, however, assume that A*B*C has very few cases and is a remainder row. Since we expect the presence of C to contribute to the outcome and we know A*B*c is sufficient, we can argue that A*B*C should obtain the outcome too. We therefore include it in the solution and then simplify to produce A*B. A remainder row like A*B*C which helps us to remove the absence of a factor we don’t expect to contribute to the outcome from the solution, is known as an ‘easy counterfactual’ (Ragin, 2008).

Suppose instead that, given a term, A*B*c, in our solution we have the remainder row A*b*c. Including this row would give the solution A*c by removing ‘B’. Our row A*b*c is acting as a ‘difficult counterfactual’ here because its inclusion into the solution amounts to removing the presence of a factor we expected, on theoretical grounds, to contribute to the outcome (Ragin, 2008). This needs more robust theoretical justification since we might expect that ‘B’ is contributing to the outcome (based on the reasoning above). Difficult counterfactuals can be incorporated into solutions but only after careful consideration.

Some of the factors in our model are not as easy to assess as those just given in the example. When considering the social class factors, for example, we could imagine that the absence (and not the presence) of the factor ‘WORKINGCLASS’ could lead to an outcome of high attainment in mathematics. Whilst we may be able to theorise that PMT-class children obtain higher mathematics test scores than working-class children, it is difficult to make judgements about the relative mathematics test scores of intermediate-class children.

Further, since I use dummy variables here to represent the three class categories, the removal of a class factor also will not necessarily reduce the number of factors in a term of the solution. For example, if we have the term, ‘MALE*PMTCLASS’, and want to include a remainder row representing intermediate-class boys in this term, we cannot simply remove the factor ‘PMTCLASS’. This would completely remove any class restriction on the term. Instead, we could introduce the factor ‘workingclass’ which does not make the term appear any simpler but does indicate a less-restrictive route to the outcome.

In the following analysis, I will attempt to construct intermediate solutions for the top 50% of mathematics attainment by, firstly, removing absent factors from any of the terms before checking the parsimonious solution to see if any of the more drastic possible simplifications can be justified. Through the use of two
models with two different crisp ability factors, I will examine the different kinds of intermediate solution that can be produced.

**Very high general ability**

The first crisp ability measure we introduce indicates whether a child is of very-high general ability. In Table 5, rows with a ‘1’ in the ‘ability (top 5%)’ column contain cases in the top 5% of general ability, as measured by the BAS. Setting such a restrictive criterion for ability has exaggerated the effect of limited diversity near the top of the table. More than half the quasi-sufficient rows in Table 5 are remainders and, hence, we expect that the most complex solution and the parsimonious solution will look very different.

**Table 5: Truth table, with the top 5% of ability, with outcome of top 50% of mathematics attainment**

| male | working -class | PMT-class | maternal interest | ability (top 5%) | number | consistency | quasi-sufficient? |
|------|----------------|-----------|-------------------|-----------------|--------|-------------|------------------|
| 1    | 0              | 0         | 0                 | 0               | 1      | 4           | yes              |
| 1    | 0              | 0         | 1                 | 1               | 15     | 1           | yes              |
| 0    | 0              | 0         | 0                 | 0               | 1      | 2           | yes              |
| 1    | 0              | 1         | 0                 | 1               | 10     | 1           | yes              |
| 1    | 0              | 1         | 1                 | 1               | 63     | 0.98        | yes              |
| 0    | 0              | 1         | 1                 | 1               | 55     | 0.95        | yes              |
| 0    | 0              | 0         | 1                 | 1               | 10     | 0.9         | yes              |
| 1    | 1              | 0         | 1                 | 1               | 26     | 0.88        | yes              |
| 0    | 0              | 1         | 0                 | 1               | 7      | 0.86        | yes              |
| 1    | 0              | 0         | 1                 | 0               | 81     | 0.83        | yes              |
| 1    | 1              | 0         | 0                 | 1               | 5      | 0.8         | yes              |
| 1    | 0              | 1         | 1                 | 1               | 281    | 0.76        | yes              |
| 0    | 1              | 0         | 1                 | 1               | 29     | 0.69        | no               |
| 0    | 0              | 1         | 1                 | 0               | 215    | 0.66        | no               |
| 0    | 0              | 0         | 1                 | 0               | 62     | 0.55        | no               |
| 1    | 0              | 1         | 0                 | 0               | 96     | 0.54        | no               |
| 1    | 1              | 0         | 1                 | 0               | 259    | 0.53        | no               |
| 0    | 1              | 0         | 1                 | 0               | 181    | 0.49        | no               |
| 1    | 0              | 0         | 0                 | 0               | 27     | 0.44        | no               |
| 0    | 0              | 1         | 0                 | 0               | 55     | 0.36        | no               |
| 0    | 0              | 0         | 0                 | 0               | 17     | 0.29        | no               |
| 1    | 1              | 0         | 0                 | 0               | 217    | 0.26        | no               |
| 0    | 1              | 0         | 0                 | 1               | 8      | 0.25        | no               |
| 0    | 1              | 0         | 0                 | 0               | 165    | 0.21        | no               |
**Creation of most-complex and parsimonious solutions**

Setting a consistency threshold of 0.75 and excluding all remainder rows gives us the solution in Figure 4. There are three routes to the outcome in this solution but only one of these routes is available to girls. Such girls would have to be in the PMT-class, top 5% of ability and have an interested mother. Also, all routes to the outcome require an interested mother and only one does not require being in the top 5% of general ability. A somewhat surprising conclusion from the solution in Figure 4 is that it is quasi-sufficient for the outcome to be an intermediate-class boy with an interested mother and not be in the top 5% of ability but not quasi-sufficient for the same type of boy who is in the top 5% of ability. This seems surprising and results like this can occur when large numbers of remainder rows are excluded.

The solution coverage figure tells us that our entire solution accounts for 39% of the outcome. Though the number of rows left out in Figure 4 is high, the resulting number of cases being excluded (which also obtain the outcome) is not. It is usual for the difference in solution coverage between the most complex and parsimonious solution to be negligible because, often, these solutions differ by only one or two rows which cover very few cases.

**Figure 4: Most-complex solution for Table 5, consistency = 0.75**

![Truth Table Solution](image)

| raw coverage | unique coverage | consistency |
|--------------|----------------|-------------|
| MALE*workingclass*MATERNAL INTEREST*ability(5%) | 0.262862 | 0.262862 | 0.776243 |
| PMTCLASS*MATERNAL INTEREST*ABILITY(5%) | 0.106642 | 0.106642 | 0.966102 |
| MALE*WORKINGCLASS*MATERNAL INTEREST*ABILITY(5%) | 0.021515 | 0.021515 | 0.884615 |

solution coverage: 0.391020
solution consistency: 0.826087

The parsimonious solution is shown in Figure 5. Because we now have included several remainder rows, we see that this parsimonious solution covers approximately 44% of the outcome as against the 39% covered by the complex solution. As in Figure 4, there are still three routes to the outcome and only one for girls. In Figure 5, however, the route for girls is less restrictive because it only requires that a girl be non-working-class and in the top 5% for ability. Working-class boys no longer need an interested mother to achieve the outcome. Figure 5 shows that it is now quasi-sufficient for them to be in the top 5% of general ability.

**Figure 5: Parsimonious solution for Table 5, consistency threshold = 0.75**

![Truth Table Solution](image)

| raw coverage | unique coverage | consistency |
|--------------|----------------|-------------|
| workingclass*ABILITY(5%) | 0.149673 | 0.073901 | 0.963855 |
| MALE*workingclass*MATERNAL INTEREST | 0.334892 | 0.262862 | 0.813636 |
| MALE*pmtclass*ABILITY(5%) | 0.043031 | 0.025257 | 0.920000 |

solution coverage: 0.437792
solution consistency: 0.837209
**Creation of intermediate solution**

We have already seen that excluding all remainder rows produces some strange conclusions but including them all can give us an over-simplified view of what is going on in the data. We need to consider each of the remainders in turn to see whether we can justify their inclusion into our simplified solution.

We can see from Table 5 that the row of 26 working-class boys of very high ability with interested mothers is quasi-sufficient (with consistency of 0.88) and we would expect intermediate-class boys of very high ability to do as well or better. We also notice that the row of intermediate-class boys with interested mothers who are not of very high ability (81 cases and a consistency of 0.83) is quasi-sufficient and we expect that boys who fit this type but are of very high ability will also achieve the outcome. Therefore, I include that row which has 15 cases and a consistency of 1. This is an example of an easy counterfactual as we are removing the factor ‘ability(5%)’ from the first term in Figure 4.

I now consider the row of intermediate-class girls of very high ability with an interested mother (10 cases, consistency 0.9). This time, the only quasi-sufficient row we can consider for comparison is also a remainder (with 2 cases and a consistency of 1). Taking a different approach, I consider that the equivalent row for boys (just discussed) has been included and using theoretical backing, I suggest that, all other factors being equal, girls are likely to achieve as well or higher in mathematics and so, I include the row of intermediate-class girls of very-high ability with interested mothers (Sammons, 1995).

The rest of the remainder rows are similar in that they all represent children who are either of very high ability or have interested mothers (apart from the 4th row from the bottom of Table 5, which has neither). It may be that children of very high ability are less in need of assistance from parents in order to do well or that a high level of maternal interest can overcome a lack of general ability but I cannot be sure enough about this to include any of these rows in the simplified solution. In summary, then, only those remainder rows representing cases in the top 5% of general ability with interested mothers are included in the intermediate solution.

The intermediate solution is shown in Figure 6. The first term now has no ability restriction (as it did in Figure 4) meaning that intermediate- or PMT-class boys with interested mothers can achieve the outcome whether they are in the top 5% of general ability or not.

We also see that ‘PMTCLASS*MATERNAL INTEREST*ABILITY(5%)’ from Figure 4 has become the less restrictive, ‘workingclass*MATERNAL INTEREST*ABILITY(5%)’ because of the inclusion of the row of intermediate-class girls of very-high ability with interested mothers. Finally, the term ‘MALE*WORKINGCLASS*MATERNAL INTEREST*ABILITY(5%)’ from Figure 4 becomes ‘MALE*pmtclass*MATERNAL INTEREST*ABILITY(5%)’ in Figure 6 because we have included the row of intermediate-class boys of very-high ability with interested mothers.

Our solution in Figure 6 does not allow for any routes to the outcome with an uninterested mother, as in Figure 5. This is because of decisions made above about which remainder rows to include and which to exclude. There is a term in Figure 5, namely ‘MALE*workingclass*MATERNAL INTEREST’, which matches one in Figure 6, however, showing that we have achieved the maximum possible degree of parsimony in that term (whilst still adhering to our aforementioned consistency threshold).

The parsimonious solution in Figure 5 allows us to make some potentially strong conclusions – namely, that for some children (even those in the working class), maternal interest is not required in order to achieve an above-average score in mathematics. We must be careful when creating solutions which include remainders that we are clear about which remainders are (and should be) included instead of striving for a solution which is the easiest to digest and which could be adopted by policymakers. Here, the intermediate solution really does lie between the most complex one and the parsimonious one but, as we shall see in the next example, this is not always the case.
Figure 6: Intermediate solution for Table 5, consistency threshold = 0.75

--- TRUTH TABLE SOLUTION ---

|                              | raw coverage | unique coverage | consistency |
|------------------------------|--------------|-----------------|-------------|
| MALE*workingclass*MATERNAL INTEREST | 0.334892     | 0.262862        | 0.813636    |
| workingclass*MATERNAL INTEREST*ABILITY(5%) | 0.129093     | 0.057063        | 0.965035    |
| MALE*pmtclass*MATERNAL INTEREST*ABILITY(5%) | 0.035547     | 0.021515        | 0.926829    |

solution coverage: 0.413471
solution consistency: 0.832392

High general ability

Our ability factor, in Table 6, now shows which children are in the top 25% for general ability. If we set a consistency threshold of 0.74, we see that there is only one quasi-sufficient counterfactual in Table 6\textsuperscript{17}.

Table 6: Truth table, with the top 25\% of ability, with outcome of top 50\% of mathematics attainment

| male | working-class | PMT-class | maternal interest | ability (25\%) | number | consistency | quasi-sufficient? |
|------|---------------|-----------|-------------------|----------------|--------|-------------|------------------|
| 0    | 1             | 1         | 1                 | 1              | 196    | 0.96        | yes              |
| 1    | 0             | 0         | 1                 | 1              | 53     | 0.94        | yes              |
| 0    | 0             | 1         | 1                 | 1              | 9      | 0.89        | yes              |
| 1    | 0             | 1         | 0                 | 1              | 40     | 0.83        | yes              |
| 0    | 0             | 0         | 1                 | 1              | 104    | 0.78        | yes              |
| 0    | 1             | 0         | 1                 | 1              | 35     | 0.77        | yes              |
| 0    | 1             | 0         | 0                 | 1              | 103    | 0.75        | yes              |
| 1    | 0             | 0         | 1                 | 0              | 43     | 0.74        | yes              |
| 0    | 0             | 0         | 0                 | 1              | 9      | 0.67        | no               |
| 1    | 0             | 1         | 1                 | 0              | 148    | 0.59        | no               |
| 0    | 0             | 1         | 0                 | 1              | 26     | 0.58        | no               |
| 1    | 1             | 0         | 0                 | 1              | 45     | 0.56        | no               |
| 0    | 0             | 1         | 1                 | 1              | 113    | 0.48        | no               |
| 1    | 0             | 1         | 0                 | 0              | 66     | 0.44        | no               |
| 1    | 1             | 0         | 1                 | 0              | 181    | 0.44        | no               |
| 0    | 0             | 0         | 1                 | 0              | 37     | 0.43        | no               |
| 0    | 1             | 0         | 0                 | 1              | 44     | 0.43        | no               |
| 0    | 0             | 1         | 0                 | 0              | 22     | 0.36        | no               |
| 0    | 0             | 0         | 1                 | 0              | 36     | 0.31        | no               |
| 0    | 1             | 0         | 0                 | 0              | 107    | 0.29        | no               |
| 1    | 1             | 0         | 0                 | 0              | 177    | 0.2         | no               |
| 0    | 1             | 0         | 0                 | 0              | 129    | 0.13        | no               |
| 0    | 0             | 0         | 0                 | 0              | 10     | 0.1         | no               |
Creation of most-complex and parsimonious solutions

As before, we start by creating the most complex version of the solution by excluding all the remainder rows in Table 6. In Figure 7, we see there are four routes to the outcome for boys and two for girls. Both the routes for girls require them to be in the top 25% for general ability and have an interested mother. A naïve user of QCA might think that combining the terms ‘pmtclass*MATERNAL INTEREST*ABILITY(25%)’ and ‘workingclass*MATERNAL INTEREST*ABILITY(25%)’ leads to the simplified term ‘INTERMEDIATECLASS*MATERNAL INTEREST*ABILITY(25%)’ but this is not actually the case.

To understand why not, we must first remember that the factors ‘pmtclass’ and ‘workingclass’ represent the absence of PMT-class and working-class cases, respectively. Hence, ‘pmtclass’ actually means ‘either INTERMEDIATECLASS or WORKINGCLASS’ and, similarly, ‘workingclass’ means ‘either INTERMEDIATECLASS or PMTCLASS’. Taking these together does not give ‘INTERMEDIATECLASS’ but, instead, ‘WORKINGCLASS or INTERMEDIATECLASS or PMTCLASS’ i.e. all possible class options. So, with a 0.74 consistency threshold, ‘MATERNAL INTEREST*ABILITY’ is quasi-sufficient for above-average attainment in mathematics, whatever the class or sex of the child.

The routes for just boys differ more starkly by class. Working-class boys must have an interested mother and be in the top 25% of general ability whilst PMT-class boys need only to be in the top 25% of general ability. Intermediate-class boys can achieve the outcome despite not being in the top 25% of general ability if they have an interested mother.

**Figure 7: Most complex solution for Table 6, consistency threshold = 0.74**

--- TRUTH TABLE SOLUTION ---

| raw coverage | unique coverage | consistency |
|--------------|----------------|-------------|
| pmtclass*MATERNAL INTEREST*ABILITY(25%) | 0.219832 | 0.147802 | 0.796610 |
| workingclass*MATERNAL INTEREST*ABILITY(25%) | 0.378859 | 0.130028 | 0.918367 |
| MALE*INTCLASS*MATERNAL INTEREST | 0.076707 | 0.029935 | 0.854167 |
| MALE*PMTCLASS*ABILITY(25%) | 0.207671 | 0.030870 | 0.940678 |

solution coverage: 0.587465
solution consistency: 0.859097

In Figure 8, we see the parsimonious solution for a 0.74 consistency threshold. Note, here, that the terms ‘pmtclass*MATERNAL INTEREST*ABILITY(25%)’ and ‘workingclass*MATERNAL INTEREST*ABILITY(25%)’ both appear again in Figure 8. The fs/QCA software will not perform the simplification discussed above because the class factors are dummy variables. This is dealt with manually in the intermediate solution below.

In fact, the term in Figure 8 which is different from Figure 7 is the term ‘MALE*workingclass*ABILITY(25%)’. This simplification arises from the inclusion of the row ‘MALE*INTCLASS*maternal interest*ABILITY(25%)’ in Table 6 (with 9 cases and a consistency of 0.89). As discussed in the example earlier, of very-high ability, we cannot be sure that this row, given more cases, would continue to exhibit a consistency which would deem it quasi-sufficient. Therefore, we do not include it in the intermediate solution. A quick check of Table 6 shows that there are no other remainder rows which we should include (based on the previously-stated criteria of having an interested mother and being in the top 5% or 25% of ability).

The intermediate solution, then, is presented in Figure 8. We discussed earlier how to combine these to produce the simplified term, ‘MATERNAL INTEREST*ABILITY(25%)’. What this output alone cannot tell
us, however, is what the consistency or coverage of that simplified term is because of the difficulties with dummy variables. We can calculate the consistency directly from Table 6 by finding all the rows with a ‘1’ in maternal interest and ability(25%), summing the number of cases which obtain the outcome and dividing by the total number of cases in these rows.

Finding the coverage is slightly more difficult. We can calculate the raw coverage for ‘MATERNAL INTEREST*ABILITY(25%)’ directly from Table 6. To find the unique coverage, we must see which of the rows represented by ‘MATERNAL INTEREST*ABILITY(25%)’ are also represented by other terms in the solution in Figure 9. The cases obtaining the outcome in these rows can be removed from the ‘MATERNAL INTEREST*ABILITY(25%)’ total and coverage re-calculated. From this calculation, we find the unique coverage of ‘MATERNAL INTEREST*ABILITY(25%)’.

**Figure 8: Parsimonious solution for Table 6, consistency 0.74**

--- TRUTH TABLE SOLUTION ---

| Term                              | raw coverage | unique coverage | consistency |
|-----------------------------------|--------------|-----------------|-------------|
| MALE*workingclass*ABILITY(25%)   | 0.261927     | 0.038354        | 0.939597    |
| pmiclass*MATERNAL INTEREST*ABILITY(25%) | 0.219832    | 0.147802        | 0.796610    |
| workingclass*MATERNAL INTEREST*ABILITY(25%) | 0.378859    | 0.130028        | 0.918367    |
| MALE*INTCLASS*MATERNAL INTEREST  | 0.076707     | 0.029934        | 0.854167    |

solution coverage: 0.594949
solution consistency: 0.859459

Figure 9 tells us that there is only one route to the outcome for girls and it is very restrictive. Girls must be in the top 25% of ability and have an interested mother. What is interesting about this term, though, is that girls of all social classes have the same route to the outcome available to them. Working-class boys, too, share this route to the outcome, i.e. they must have an interested mother and be in the top 25% of ability to achieve the outcome. Boys of the intermediate-class can achieve the outcome if they have an interested mother, regardless of their ability and boys of PMT-class can achieve the outcome if they are in the top 25% of ability even without an interested mother. This may seem surprising, as we may expect an easier route to the outcome for children of the highest social classes. What we see here is that class does not make a highly able girl’s route to the outcome any less or more restrictive and, for boys, those in the highest social class, the PMT-class, still must be in the top 25% of ability to achieve the outcome. For intermediate-class boys, however, their route to the outcome does rely on a factor, maternal interest, over which they have limited control.

These results begin to give an insight into the possible effects of parental involvement in education on mathematics attainment. What we cannot pick out here, of course, is effects due to any differences in type of involvement by class though we might begin to hypothesise such differences could be at the root of these class differences in the effects of involvement.
Figure 9: Intermediate solution for Table 6, consistency threshold = 0.74

--- TRUTH TABLE SOLUTION ---

|                | raw coverage | unique coverage | consistency  |
|----------------|--------------|-----------------|--------------|
| MALE*INTCLASS*MATERNAL INTEREST | 0.076707     | 0.029935        | 0.854167     |
| MATERNAL INTEREST*ABILITY(25%)    | 0.526660     | 0.277830        | 0.868827     |
| MALE*PMTCLASS*ABILITY(25%)        | 0.207671     | 0.030870        | 0.940678     |

solution coverage: 0.587465
solution consistency: 0.859097

Conclusion

Previous work on parental involvement in education has either assumed that parental involvement will produce ‘unique’ effects on attainment (independent of any other factors) or has assumed that any effects of parental involvement will be necessarily intertwined with contextual factors and tried to describe, in detail what these contextual factors are and why they may modify the effect of involvement (Desforges, 2003). As a result, we find that there is a large body of empirical work which lacks detail about the nature of any configurational causal processes at work and a similarly vast body of qualitative literature which relies on less systematic approaches to analysis.

I have used the language and method of QCA to discuss parental involvement in terms of sufficient configurations for mathematics attainment. Re-framing a discussion about the conditional effects of parental involvement on mathematics attainment in this way allows me to uncover some of the complex causal processes at work. I can look for, as I have here, quasi-sufficient configurations of sex, social class, maternal interest and ability and comment on the degree of sufficiency and degree of coverage of solutions.

My initial analysis, not presented here, showed that there were no simplified configurations which were quasi-sufficient for very-high attainment or high attainment in mathematics. Relaxing the outcome measure to account for above-average mathematics attainment did yield configurations which were quasi-sufficient for attainment but created truth tables where most rows were closer, in consistency, to 0.5 than to 0 or 1. Rows such as these could be said to be neither quasi-sufficient for the outcome nor quasi-sufficient for its negation.

I then refined the model by introducing a new factor – general ability. Analysis of the refined model produced truth tables where some remainder rows had very-high consistencies but such a low number of cases that we could not confidently say their consistency score was a reflection of the sufficiency of the configuration. Similarly, some remainder rows had very-low, potentially un-trustworthy, consistency scores. I treat all these remainder rows as though they contain no cases and make a theoretical decision about whether they can be included in the solution or not. I, in each case, create two solutions, the most complex and parsimonious, to act as boundaries for complexity and move to create an intermediate solution which only includes those counterfactuals which I, theoretically, expect to obtain the outcome. Throughout each stage of the analysis, we see that girls face a more restrictive route to the outcome of mathematics attainment as they must, typically, be in a higher social class and/or be of the highest general ability level with an interested mother.

Notes

1 I would like to thank the anonymous reviewers and the Special Issue Editors for their many helpful comments on an earlier draft of this paper.
In the field of parental involvement in mathematics, we could imagine that a working-class boy with an interested mother and a working-class boy with an uninterested mother get the same high score on a mathematics test. Despite their identical test marks, in a case-based approach, we are free to interpret these as two very different results if we suspect a theoretical link between high parental interest and test scores.

In QCA, configurations of 'factors' constitute a case and a case is not decomposed into these at any point in the analysis. The focus is on how cases in their entirety behave, though the effect of changing one condition on the outcome experienced by a case comprising an otherwise unchanged configuration of factors might be an important focus in some contexts.

It should be noted that some authors (e.g. Schneider and Wagemann, 2010) suggest that all QCA analyses should include, and indeed start with, analyses of necessary conditions.

Ragin (2000) draws a distinction between a 'veristic' approach to assessing sufficiency and a probabilistic one. In the veristic approach, we can call a configuration ‘perfectly sufficient’ if all cases (however small the number) of that configuration obtain the outcome. When low numbers of cases yield near-perfect consistencies, however, we must be willing to take into account that the cases we have happen, by chance, to be the ones of that type which do obtain the outcome.

Ragin (2008) suggests a lower limit of consistency of 0.75 as it is difficult to claim a genuine subset relation if the row is less consistent than this. In our analyses, we aim to adhere to this lower limit but discuss whether, in each instance, rows which fall beneath this level can be legitimately excluded.

A consistency of 0.8 or above, for example, can tell us that a particular configuration is ‘almost always’ sufficient whilst 0.7 could indicate the configuration is ‘usually sufficient’ (Ragin, 2000).

In this table and all others in the paper, consistency scores are given to 2 decimal places.

The other rows covered by b*D are a*b*c*D, A*b*C*D and a*b*C*D. All three are also included in/covered by the other terms in the solution, a*D and C*D.

A separate analysis was conducted for father’s interest in the child’s education but the results are not included here.

These 1662 cases are those without missing data in any of the variables used to construct the factors. The whole-sample mean test scores and standard deviations for the mathematics attainment variable and general ability variable are lower than those in this smaller sample and we have, proportionally, a higher number of PMT-class children (36% instead of 29%) and a lower proportion of working-class children (52% instead of 60%) in this sample. Our smaller sample also has a higher proportion (58% instead of 52%) of boys than BCS70.

For more on the sociological problems with the RG scheme, see Prandy (1999).

In the truth table generated by the fs/QCA software, there is a column (headed by the name of the outcome measure) in which the researcher inputs 1’s and 0’s to indicate which rows are in the solution.

Rows with no cases at all will have a consistency of 0 because we are working with crisp-sets.

Here, each remainder row is treated as either consistent with sufficiency or not purely on the basis of which choice makes for the most parsimonious solution.

If we include factors whose absence is thought to contribute to Y, the reasoning that follows would be reversed for those factors.

I include the 9th row here because it has only a negligible difference in consistency from the 8th row and is very close to the threshold we want to impose of 0.75. When setting a consistency threshold, we should be aware of rows which may have similar consistencies but will be split on either side of the threshold by imposing that threshold too mechanically.

References

Bakker, J., Denessen, E., & Brus-Laeven, M. (2007). Socio-Economic Background, Parental Involvement and Teacher Perceptions of these in relation to Pupil Achievement. Educational Studies, 33(2), 177-192.

Cooper, B. (2005). Applying Ragin’s Crisp and Fuzzy Set QCA to Large Datasets: Social Class and Educational Achievement in the National Child Development Study. Sociological Research Online, 10(2).

Cooper, B., & Glaesser, J. (2008). How Has Educational Expansion Changed the Necessary and Sufficient Conditions for Achieving Professional, Managerial and Technical Class Positions in Britain? A Configurational Analysis. Sociological Research Online, 13(3), 2.
Cooper, B., & Glaesser, J. (2009). *Using Case-Based Approaches to Analyse Large Datasets - A Comparison of Ragin's fsQCA and Cluster Analysis.* Paper presented at the British Sociological Association Annual Conference.

Crozier, G. (1999). Is it a case of “We know when we’re not wanted?” The parents’ perspective on parent-teacher roles and relationships. *Educational Research, 41*(3), 315-328.

Desforges, C. (2003). *The Impact of Parental Involvement, Parental Support and Family Education on Pupil Achievements and Adjustment: A Literature Review.* London: DfES.

Domina, T. (2005). Levelling the Home Advantage: Assessing the Effectiveness of Parental Involvement in Elementary School. *Sociology of Education, 78*, 233-249.

Friedel, J. M., Cortina, K. S., Turner, J. C., & Midgley, C. (2007). Achievement Goals, Efficacy Beliefs and Coping Strategies in Mathematics: The Roles of Perceived Parent and Teacher Goal Emphases. *Contemporary Educational Psychology, 32*, 434-458.

Hirata, J., Nishimura, K., Urasaka, J., & Yagi, T. (2006). Parents' Educational Background, Subjects "Good-At" in School and Income: An Empirical Study. *The Japanese Economic Review, 57*(5), 533-546.

Mahoney, J., & Goertz, G. (2006). A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research. *Political Analysis, 14*(3), 227-249.

Prandy, K. (1999). Class, stratification and inequalities in health: a comparison of the Registrar-General’s Social Classes and the Cambridge Scale. *Sociology of Health & Illness, 21*(4), 466-484.

Ragin, C. C. (1987). *The Comparative Method. Moving beyond Qualitative and Quantitative Strategies.* Berkeley: University of California Press.

Ragin C. C. (2000). *Fuzzy-set social science.* Chicago: University of Chicago Press.

Ragin, C.C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond.* Chicago: University of Chicago Press.

Reay, D. (1998). *Class Work: Mothers' Involvement in their Children's Primary Schooling.* London: University College Press.

Sammons, P. (1995). Gender, Ethnic and Socio-economic Differences in Attainment and Progress: a longitudinal analysis of student achievement over 9 years. *British Educational Research Journal, 21*(4), 465-485.

Schneider, C. Q. and Wagemann, C. (2010). Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-Sets. *Comparative Sociology, 9*, 1-22.

**Biography**

Stephanie Thomson is a doctoral researcher in the School of Education, Durham University, UK. She is currently researching parental involvement in primary mathematics.