UCL Centre for Education Policy & Equalising Opportunities

CEPEO Working Paper No. 26-08

Exhausting exams: the wellbeing effects of compulsory retakes

Robbie Maris
UCL and EPI

Jake Anders
UCL

Tammy Campbell
EPI

Gill Wyness
UCL

Policies that require low-attaining students to continue investing in foundational skills may generate future human-capital benefits but also impose short-run psychological costs. We study this trade-off using England's compulsory GCSE English and mathematics resit requirement. Using longitudinal survey data from the Millennium Cohort Study (MCS), we estimate the effect of exposure to England's compulsory resit requirement on adolescent wellbeing, controlling for cognitive ability, prior wellbeing and a rich set of demographic and socioeconomic characteristics. We conduct a cross-country falsification test using Wales and Northern Ireland, where students also take GCSEs but face no comparable national resit requirement. We find that exposure to the requirement lowers wellbeing at age 17 by approximately 0.11 SDs. These effects are about exposure to retakes at the extensive margin - we find no evidence of an additional penalty for retaking in multiple subjects. Our average estimates conceal substantial heterogeneity: English resitting is associated with lower wellbeing among boys, while mathematics resitting is associated with lower wellbeing among girls. For boys resitting English, the negative estimates are concentrated among students further below the pass threshold, consistent with a discouragement mechanism; for girls resitting mathematics, they are concentrated among those who narrowly miss the threshold, consistent with a near-miss penalty. The results identify a potentially important non-pecuniary cost of compulsory remedial education and highlight a trade-off between policies designed to raise foundational skills and adolescent wellbeing.

VERSION: July 2026

Suggested citation: Maris, R, Anders, J., Campbell, T. & Wyness, G. (2026). *Exhausting Exams: The Wellbeing Effects of Compulsory Retakes.* (CEPEO Working Paper No. 26-08). UCL Centre for Education Policy and Equalising Opportunities. https://econpapers.repec.org/RePEc:ucl:cepeow:26-08.

Disclaimer

Any opinions expressed here are those of the author(s) and not those of UCL. Research published in this series may include views on policy, but the university itself takes no institutional policy positions.

CEPEO Working Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Highlights

- We use nationally representative Millennium Cohort Study data to examine the wellbeing effects of England's compulsory post-16 English and maths resit policy.
- Students required to continue studying English and/or maths after failing to achieve a standard GCSE pass have wellbeing around 0.11 standard deviations lower at age 17, after accounting for prior wellbeing, cognitive ability and a rich set of characteristics.
- The wellbeing penalty appears to be linked to whether or not a student has to do any resits: we find no evidence of an additional penalty for students required to resit both English and maths rather than one subject.
- Effects vary substantially across students and subjects. Negative effects are concentrated among boys resitting English and girls resitting maths, with further differences according to how far students were from the original pass threshold.
- As a falsification test, we examine students falling below the same GCSE thresholds in Wales and Northern Ireland, where there is no equivalent national resit requirement. The corresponding effect on wellbeing is close to zero.

Why does this matter?

Ensuring young people develop strong English and maths skills is important, but compulsory resits may come with meaningful wellbeing costs. Designing post-16 support around students' needs and readiness could help build essential skills without unnecessarily reinforcing the experience of failure.

Exhausting Exams: The Wellbeing Effects of Compulsory Retakes

Robbie Maris* Tammy Campbell† Jake Anders‡ Gill Wyness§

August 18, 2026

Abstract

Policies that require low-attaining students to continue investing in foundational skills may generate future human-capital benefits but also impose short-run psychological costs. We study this trade-off using England's compulsory GCSE English and mathematics resit requirement. Using longitudinal survey data from the Millennium Cohort Study (MCS), we estimate the effect of exposure to England's compulsory resit requirement on adolescent wellbeing, controlling for cognitive ability, prior wellbeing and a rich set of demographic and socioeconomic characteristics. We conduct a cross-country falsification test using Wales and Northern Ireland, where students also take GCSEs but face no comparable national resit requirement. We find that exposure to the requirement lowers wellbeing at age 17 by approximately 0.11 SDs. These effects are about exposure to retakes at the extensive margin – we find no evidence of an additional penalty for retaking in multiple subjects. Our average estimates conceal substantial heterogeneity: English resitting is associated with lower wellbeing among boys, while mathematics resitting is associated with lower wellbeing among girls. For boys resitting English, the negative estimates are concentrated among students further below the pass threshold, consistent with a discouragement mechanism; for girls resitting mathematics, they are concentrated among those who narrowly miss the threshold, consistent with a near-miss penalty. The results identify a potentially important non-pecuniary cost of compulsory remedial education and highlight a trade-off between policies designed to raise foundational skills and adolescent wellbeing.

*UCL, CEPEO and Education Policy Institute. Email: robert.maris.23@ucl.ac.uk.
†Education Policy Institute and CASE, LSE. Email: tammy.campbell@epi.org.uk.
‡UCL, CEPEO. Email: jake.anders@ucl.ac.uk.
§UCL CEPEO and Centre for Economic Performance, LSE. Email: g.wyness@ucl.ac.uk.

1 Introduction

Governments frequently use compulsory education and remediation requirements to increase human-capital investment among individuals who have not reached minimum skill standards (Angrist and Krueger, 1991; Devereux and Hart, 2010; Jacob and Lefgren, 2004). A large literature shows that compulsory schooling can increase educational attainment and generate subsequent economic returns (Angrist and Krueger, 1991). Yet, additional compulsory study is not costless. Remediation and repetition require further time and effort and can affect subsequent educational participation (Jacob and Lefgren, 2009; Tafreschi and Thiemann, 2016), while repeated attempts at high-stakes examinations can impose private costs on students (Krishna et al., 2018). Despite increasing interest in the relationship between education and mental health, much less is known about the short-run wellbeing consequences of compelling low-attaining adolescents to continue studying and being assessed in subjects in which they have already experienced failure (Avendano et al., 2020).

"Sucking the life out of students" (Denny, 2026). This is one description of England's compulsory resit policy. Under the policy, students who do not achieve a standard pass in GCSE English or mathematics at age 16 are required to continue studying the relevant subject during post-16 education. The policy is intended to strengthen young people's literacy and numeracy and improve their subsequent opportunities. However, requiring students to return repeatedly to subjects and assessments in which they have already struggled may also impose costs on confidence, motivation and mental wellbeing.

This paper asks whether exposure to the compulsory English and mathematics resit requirement lowers adolescent wellbeing. We use this setting to estimate the short-run wellbeing cost of compulsory remediation following academic failure. We also provide the first estimates of the effects of England's national resit policy on mental wellbeing. Previous work shows high-stakes examinations and examination failure can adversely affect student wellbeing (Beck et al., 2024; Fishstrom et al., 2022; Jerrim, 2023). However, little is known about the additional wellbeing consequences of compulsory study and remediation following failure. England's GCSE resit requirement provides a setting in which to examine this distinct margin.

To analyse the impacts of resits on wellbeing, we use data from the Millennium Cohort Study (MCS), which is a longitudinal cohort study following students born between 2000 and 2002 in the UK. This dataset is useful as the cohort members in England would have been finishing their GCSEs in 2016/17 when the resit policy was in effect (introduced in 2014/15).

We identify students who were required to resit GCSE English and/or maths and com-

pare their wellbeing outcomes to otherwise similar students who were not required to resit. To address differences between groups, we control for cognitive ability using literacy and numeracy assessments collected close to the resit period, alongside a rich set of demographic and socioeconomic characteristics. We also control for prior wellbeing, reducing the scope for reverse causality and omitted variable bias arising from persistent differences in underlying mental health. We supplement this approach with a cross-country falsification test using Wales and Northern Ireland, where students could fail English and mathematics GCSEs but were not subject to the equivalent national resit requirement.

We find that exposure to compulsory resits lowers adolescent wellbeing. We estimate that exposure to at least one resit requirement lowers mental wellbeing by 0.11 SDs. We show that this effect is about exposure to retake requirements at the extensive margin – being required to resit multiple subjects has no additional effect on wellbeing.

The average estimates conceal substantial heterogeneity. The negative English effect is concentrated among boys, particularly those who are further below the pass threshold. The English point estimate is also larger among students with an Education, Health and Care Plan (EHCP), although this estimate is imprecise. Mathematics resit status is linked to lower wellbeing among girls, particularly those close to the passing threshold, and among respondents interviewed in January or February. The item-level estimates suggest that the relationships are concentrated in particular aspects of wellbeing (feeling useful and thinking clearly), rather than appearing uniformly across all dimensions of wellbeing.

Our falsification tests, which involve replicating the analysis in Wales and Northern Ireland, do not reproduce the England patterns. This provides support for our identification strategy and suggests there is an impact of resitting beyond the initial exam failure itself. We conduct a number of other robustness checks and our results remain stable.

This paper contributes to three related literatures. First, a substantial and growing literature examines the psychological consequences of education and high-stakes assessment. Research on test anxiety shows that the pressure associated with high-stakes exams is associated with adverse educational and psychological outcomes, while recent quasi-experimental evidence indicates that exposure to national high-stakes testing can increase school-related stress among adolescents (Beck et al., 2024; Högberg and Horn, 2022; Jerrim, 2023; von der Embse et al., 2018). More broadly, a growing economics literature examines the relationship between education and mental health, including the effects of compulsory schooling (Avendano et al., 2020). We add to this literature by studying a different margin of compulsory education policy: the short-run wellbeing consequences of requiring low-attaining adolescents to continue investing in skills after academic failure. To our knowledge, we provide the first empirical estimates of the wellbeing consequences of compulsory exam retakes.

Second, we contribute to research on academic failure, remediation and repeated assessment. Existing studies show that narrowly missing important qualification thresholds can have substantial consequences for subsequent educational participation and progression (Machin et al., 2020), while failing a high-stakes examination may also affect mental health (Beck et al., 2024). A related economics literature examines policy responses to low achievement, including remedial education, grade retention and repeated attempts at high-stakes examinations (Jacob and Lefgren, 2004, 2009; Krishna et al., 2018). This work has primarily focused on attainment, progression and the incentives created by opportunities to retake. We instead examine a less studied cost margin: the wellbeing consequences of being required to continue studying and being reassessed after initial failure. This distinction is particularly important in England, where students below the GCSE English or mathematics pass threshold must generally continue studying these subjects during post-16 education.

Finally, we add to the small pool of existing research on GCSE resits, which has focused mainly on attainment, progression and the organisation of resit provision (Maris and Robinson, 2025). However, some qualitative evidence identifies low confidence, feelings of failure, anxiety and disengagement as important barriers among resit students (Crisp et al., 2023; Maris and Robinson, 2025). We extend the literature by providing the first quantitative analysis, to our knowledge, of the relationship between England's compulsory GCSE resit policy and adolescent mental wellbeing. We also examine heterogeneity by subject, gender and dimensions of wellbeing, helping to identify which students and which aspects of wellbeing may be most affected.

In the following section, we provide relevant background information on the English education system and the resit policy. We then discuss the data we use and our methods and finish by presenting results alongside a series of robustness checks. We conclude with recommendations for future research and policy.

2 Background

This section introduces the English education system and England's resit policy to provide background for the remainder of the paper.

2.1 England's education system

In England, education is compulsory until the age of 18 (the participation age was gradually raised from 16 to 18 between 2013 and 2015). The education system is organised into five key stages (KS). Of these, KS4 and KS5 are the most relevant to this paper. KS4 covers

years 10 and 11 where students are aged between 14 and 16. During this period, students study towards their GCSEs, which are predominantly examined at the end of KS4 (aged 16) in a set of high-stakes exams (Machin et al., 2020). English and maths are two of the three mandatory GCSE subjects (science being the other). Students at KS4 are primarily studying towards level 2 awards and qualifications (a standard pass in a GCSE is a level 2 award).

KS5 covers the ages of 16 through to 19 (upper-secondary, also called post-16) where most students either study for two or three years (till the age of 18 or 19).[1] The majority of students study level 3 qualifications in this phase (Tuckett, 2025) but many also study towards qualifications at level 2 or lower (this includes GCSE retakes). Level 3 qualifications include A levels, vocational/technical qualifications (like BTECs and the new T levels) and apprenticeships (Maris et al., 2024).

Students undertake KS4 in lower secondary schools and then usually attend either a further education (FE) college, sixth form college or school sixth form (lower secondary school with a sixth form attached). Students at sixth form colleges and school sixth forms are much more likely to be studying towards level 3 academic qualifications (A levels). On the other hand, students at FE colleges take a more diverse range of qualifications and levels.

2.2 England's resit policy

The resit policy, formally known as the condition of funding, requires students who do not achieve a standard pass in their GCSE English and maths to continue studying these subjects during their upper-secondary (16 to 18) education (Education and Skills Funding Agency and Department for Education, 2025). The policy was introduced in 2014 and one of the core aims was to improve literacy and numeracy skills among young adults. Prior to its introduction, young adults (16 to 24 year olds) in England had one of the lowest levels of literacy and numeracy skills across the entire OECD (OECD, 2013).

Students who achieve a grade 3 (or grade D) in their GCSEs are required to continue studying for the GCSE exam. Students with lower grades are allowed to study towards a pass in a level 2 functional skills qualification (FSQ) and are allowed to enter lower-level stepping stone qualifications (which are usually entry level or level 1 FSQs) when working towards this end goal. In practice, however, a large majority of students (80 to 85% in 2023) sit GCSE exams rather than FSQs (Maris, 2024). This is for a number of reasons, including that GCSEs have greater currency with stakeholders (parents, students, teachers

[1] Almost all qualifications and programmes are between one and two years in length. Reasons for studying for a third year vary but could be because the student studied a smaller or lower level qualification in their first year and then moved on to a two year qualification in the second year.

and employers) and they have a graduated grading scale rather than a binary pass or fail (Association of Colleges, 2022; Maris and Robinson, 2025; The Royal Society, 2021).

Students tend to sit these exams in the standard Summer exam window (May to June). However, students can also be entered for November resits which occur at the start of the academic year (in November). Both sets of exam windows are graded to the same standards and November resits are more common for maths than English (Maris and Robinson, 2025).

In general, it is widely acknowledged that supporting students to further develop their literacy and numeracy skills, particularly students with lower prior attainment, during post-16 education is an important goal. Indeed, in the recent Curriculum and Assessment Review, the resit policy was retained with some changes designed to improve delivery and outcomes. However, the delivery of the resit policy to date has been criticised on a number of occasions.

2.3 Criticisms and potential impacts on wellbeing

The resit policy has also been relatively controversial and sparked much discussion and debate in the sector. One only needs to look at sector news outlets (like FE Week) to gauge the extent of the debate around English and maths resits. Many have criticised the policy for placing significant additional costs on an already stretched FE sector and for what has been described as the *"demoralising"* nature of forcing students to resit the same exams over and over again (AOC, 2023; Camden, 2023; Monk, 2025). Others have described resits as a *"remorseless treadmill"* and as *"sucking the life out of students"* (Busby, 2024; Denny, 2026). This provides clear motivation for analysing the impacts of the resit policy on student wellbeing.

3 Conceptual framework

Here we present a brief conceptual framework showing why we might expect compulsory retakes to impact wellbeing. It also allows us to describe where we might expect the impacts to be the greatest.

3.1 Failure, resitting and wellbeing

To begin, it is worth noting that failing GCSE English or maths could affect wellbeing through two distinct channels: a) the costs associated with perceived failure and b) the costs associated with being required to resit. Examination failure may itself generate disappointment, reduce academic confidence and restrict access to subsequent education or employment

(Beck et al., 2024). Second, conditional on initial failure, compulsory resitting may impose additional benefits and costs (the focus of this paper).

Let F_{is} indicate that student i fails to achieve a standard pass in subject s, and let R_{isc} indicate exposure to a compulsory resit requirement in country c. The contemporaneous change in wellbeing can be represented as:

$$\Delta W_{isc} = -\lambda_{is} F_{is} + R_{isc} \left(p_{is} V_{is} - C_{is} \right), \tag{1}$$

where λ_{is} captures the contemporaneous cost of initial failure, p_{is} is the student's perceived probability of eventually passing, V_{is} is the perceived value of obtaining a pass, and C_{is} captures the immediate psychological, effort and opportunity costs of continued study and examination.

A resit provides another opportunity to obtain an important qualification and may improve later educational and labour-market opportunities (Bizopoulou et al., 2024; Goodman et al., 2020). However, compulsory study may also generate examination anxiety, displace time from the student's principal programme and reinforce feelings of failure or low competence (Crisp et al., 2023; Maris and Robinson, 2025). Resitting therefore lowers contemporaneous wellbeing where:

$$p_{is} V_{is} < C_{is}. \tag{2}$$

This does not imply that resitting necessarily lowers longer-term welfare: passing may generate substantial future benefits. Rather, the framework describes when the perceived future benefit is insufficient to compensate for the immediate cost at the point when wellbeing is measured.

The distinction between failure and compulsory resitting is central to the empirical analysis. In England, falling below the English or mathematics pass threshold generally triggers continued compulsory study. The observed England association may therefore combine the effect of initial failure with the additional consequences of the resit mandate. In Wales and Northern Ireland, the same national requirement did not apply to this cohort. Comparing England estimates to Wales and Northern Ireland will provide a test of whether patterns can be explained largely by initial exam failure rather than the resit requirements.

3.2 Sources of heterogeneity

The terms in Equation 1 are likely to vary across students.

First, the perceived probability of passing should generally be lower among students further below the threshold. Compulsory study may consequently appear less valuable and

more discouraging for these students. There is nevertheless a competing near-miss mechanism: narrowly failing may generate particularly salient disappointment because a passing outcome was readily imaginable (Roese, 1997). The relationship between distance from the threshold and wellbeing is therefore not necessarily monotonic.

Second, effects may differ by gender and subject. Boys tend to report weaker English or verbal self-concept, while girls tend to report weaker mathematics self-concept and greater mathematics anxiety (Devine et al., 2012; Nagy et al., 2010; Sullivan, 2009). An initial failure and compulsory resit may therefore reinforce subject-specific beliefs about low competence. We consequently expect English resits to be more consequential for boys and mathematics resits to be more consequential for girls.

Third, students with special educational needs and disabilities may face additional barriers to repeated study and high-stakes assessment. For example, students with specific learning difficulties may face persistent barriers in reading, writing or numeracy, while students with attention, communication or emotional and behavioural needs may find repeated classroom engagement and high-stakes examination preparation especially demanding. Repeated exposure to an assessment format in which a student has previously struggled may reinforce low academic confidence or feelings of failure. Young people with learning or neurodevelopmental difficulties also experience poorer mental-health outcomes on average, making this group potentially more vulnerable to additional academic stress (Finning et al., 2022; Hatton et al., 2018).

Finally, the psychological cost of resitting may vary over the academic year. If examination anxiety and revision pressure are important components of C_{is}, associations should be most negative near the summer examination period. Later estimates may additionally reflect post-examination uncertainty, cumulative exposure and selection among students who remain below the threshold. Likewise, if there is a wellbeing cost to initial exam failure (i.e., $\lambda_{is} > 0$), we would expect this to be larger closer to when the student finds out about the failure and to dissipate over time. For example, Beck et al. (2024) find that failing a high-stakes examination is associated with an increase in psychological diagnoses during the following year, but not during the second year after the examination.

A further source of heterogeneity lies in the different dimensions of wellbeing captured by the outcome scale. Composite wellbeing scales aggregate responses across several domains, but compulsory resitting may not affect each domain equally. In this paper, we use the shortened Warwick-Edinburgh Wellbeing Scale which consists of seven distinct items (Tennant et al., 2007). We would expect to see the largest effects for items that are more closely linked to the benefits and costs of resitting (V_{is} and C_{is}). Repeated academic failure and continued compulsory study may reduce feelings of usefulness and optimism about the

future, while having much weaker effects on domains such as social connectedness or general decision-making. Items relating to optimism and feeling useful are closely connected to perceived competence, self-efficacy and positive expectations about the future (Avedissian and Alayan, 2021). We therefore expect the aggregate wellbeing effects to be concentrated in a subset of theoretically relevant items rather than distributed uniformly across the scale.

3.3 Empirical implications

The framework motivates five empirical implications:

1. Effects should be strongest for wellbeing items closely related to competence and future expectations;
2. Effects may differ by gender and subject. English resits may be particularly consequential for boys, while mathematics resits may be particularly consequential for girls, reflecting differences in subject-specific self-concept, anxiety and gendered academic identities.
3. Effects may vary with distance from the pass threshold through competing discouragement and near-miss mechanisms;
4. Effects may differ by SEND status and proximity to the summer examination period; and
5. If there is a resitting effect (rather than just an examination failure effect), we would expect stronger negative associations in England than Wales and Northern Ireland (where no comparable resit requirement exists).

Our data do not directly measure perceived pass probabilities, subject-specific self-concept, examination anxiety or the support received during resit provision. Nonetheless, the subgroup analyses are interpreted as theory-guided heterogeneity and indicators of mechanisms, rather than direct tests of effect mediation.

4 Data

To examine the relationship between resits and wellbeing, we use data from the Millennium Cohort Study (MCS), a nationally representative longitudinal study of approximately 19,000 children born in the UK between September 2000 and January 2002 (Joshi and Fitzsimons,

2016). Participants have been followed from infancy through repeated survey sweeps, collecting information on educational attainment, cognitive development, family background, health, and wellbeing.

We primarily use data from Sweep 6 (age 14) and Sweep 7 (age 17). Sweep 6 includes literacy cognitive assessments, wellbeing and family background measures that are collected before resits take place. Sweep 7 includes numeracy cognitive ability, wellbeing measures and information on GCSE grades (from exams at age 16). Sweep 7 in England was collected while respondents were 17, meaning they were in the first year of their 16-18 education (year 12). This represents an optimal time to measure the impact of doing resits on wellbeing.

We only include students in England who are not missing data for our key outcome and explanatory variables. The maximum sample size is N = 4,776, while the fully controlled specification uses N = 4,358.

4.1 Identifying resit students

In Sweep 7 of the MCS, cohort members are asked about all the qualifications they hold and the grades they received. From this, we can derive whether a student fell short of passing their GCSE English Language and GCSE mathematics qualifications and therefore have to resit under the condition of funding.

It should be worth noting two limitations. First, this is measured partway through 16 to 18 education (halfway through year 12). If a resit student passes before they were surveyed, they will likely report a passing grade and will not be identified as a resit student. Students can resit in November and the traditional summer exam series (May–June).

As Figure 1 shows, however, 97% of the England MCS sample were interviewed by the end of August 2018, before the beginning of the subsequent academic year. This means most students would not have received results from summer resits (if they sat them), which were released towards the end of August.

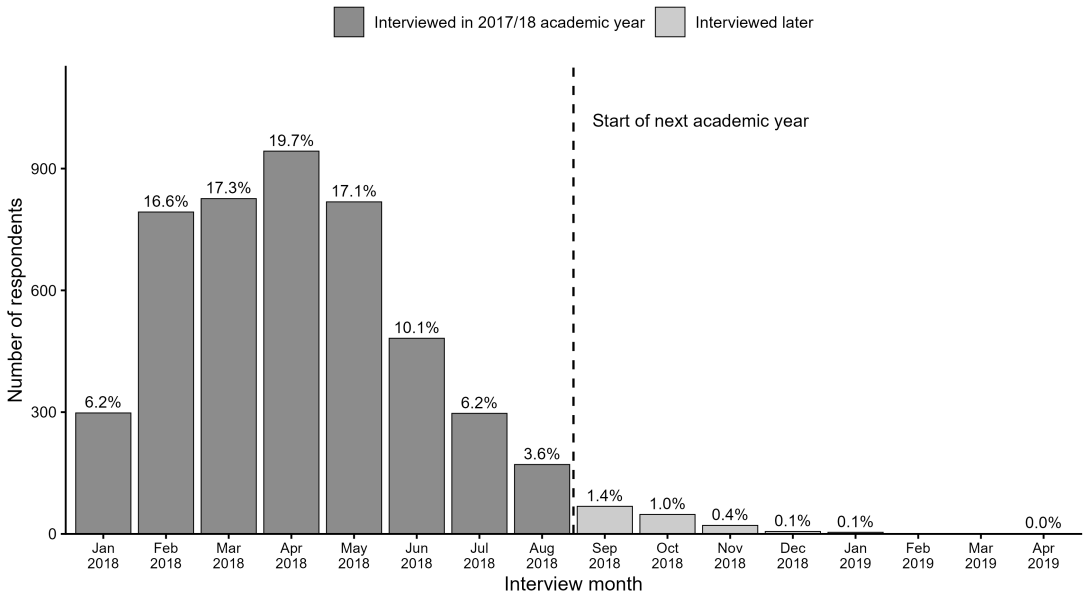


Figure 1: Timing of MCS Sweep 7 interviews in the England analysis sample

As a robustness check, we exclude the 3% of respondents who were interviewed after August and our results remain stable. Throughout all of our analyses, we additionally control for the timing of the interview.

November resits could potentially be a larger concern if many students are entered for, and passed, a November resit (Maris and Robinson, 2025). Using administrative data from the National Pupil Database (NPD), we checked what proportion of resit–eligible students fell into this category in the relevant period (2017/18 November exams). We found that only 2.1% and 3.6% of English and maths students successfully passed a November resit. Therefore, we do not see this as major limitation of our identification of resit students.

Our second potential concern is that we are using self-reported grades. Stopforth et al. (2025) showed there was some inconsistency between self-reported grades in the MCS and grades recorded in the administrative data. There was some measurement error in both the positive and negative direction (and missingness), but their example revealed this had little impact on their empirical results (Stopforth et al., 2025). In the Appendix, we consider these implications in more detail.

From these data, we identify that **18.4% of students** are required to be resitting English and maths at the time the MCS is collected (N = 880). Of these students, **42% failed both** maths and English, 30% failed only maths and 28% failed only English.

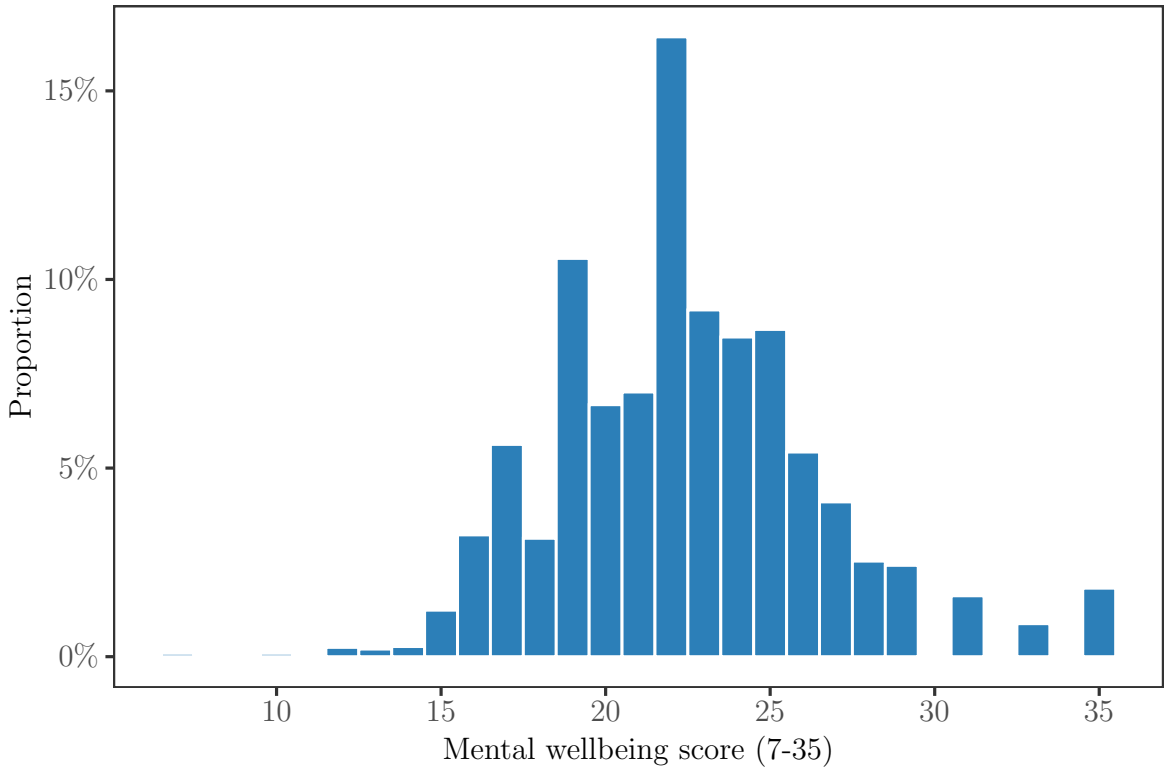
4.2 Wellbeing measure

In the MCS Sweep 7 (age 17), individuals are asked questions from the Warwick-Edinburgh Mental Wellbeing Scale (Tennant et al., 2007). The scale includes seven questions coded on a Likert scale from 1 to 5 (1 = none of the time, 5 = all of the time). Students are asked how often during the past two weeks they have been:

1. feeling optimistic about the future.
2. feeling useful
3. feeling relaxed
4. dealing with problems well
5. thinking clearly
6. feeling close to other people
7. able to make my own mind about things

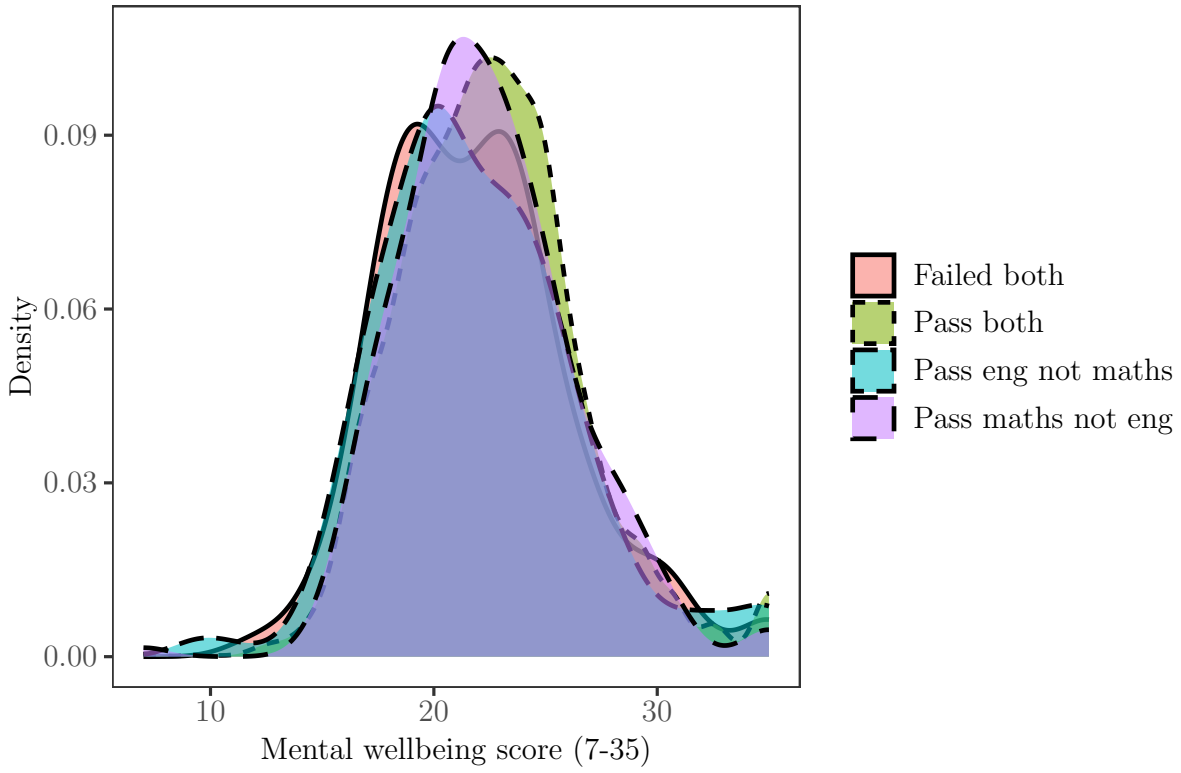
The scale ranges from 7 to 35. It is relatively normally distributed with a median of 22 (see Figure 2). Individual items are shown in the Appendix in Figure B1.

Figure 2: Wellbeing distribution



When we look at the wellbeing distribution by resit group (Figure 3), the distributions overlap significantly. However, it appears that those who pass both subjects may have slightly higher wellbeing on average.

Figure 3: Density plots of wellbeing by resit group



4.2.1 Prior wellbeing control

In Sweep 6 (age 14), cohort members complete a wellbeing grid capturing subjective wellbeing at age 14. Cohort members are asked: *"On a scale of 1 to 7 where '1' means completely happy and '7' means not at all happy, how do you feel about the following parts of your life?"*

1. Your school work?
2. The way you look?
3. Your family?
4. Your friends?
5. The school you go to?
6. Life as a whole?

We use this wellbeing index as a control for prior wellbeing. This scale is measured on a 6 to 42 scale and is standardised before we include it in our modelling. This allows us to examine how doing resits shifts wellbeing among young people.

4.3 Cognitive ability controls

In the MCS, cohort members are regularly asked to complete tasks and assessments to measure cognitive ability. These are useful in controlling for underlying ability when we are interested in either educational outcomes or education-related interventions or treatments (for example, having to resit English and maths). These will help reduce the potential confounding that ability could introduce when analysing the relationship between resits and wellbeing.

In general, cognitive ability measures are perceived to be relatively stable over time, particularly when compared with standard test scores (Breit et al., 2024; Gustavson et al., 2025). However, research has shown that the stability of these measures increases rapidly throughout childhood. Measures of cognitive ability in early childhood show only moderate levels of stability, while in adolescence they show very strong levels of stability (Breit et al., 2024; Gustavson et al., 2025).

For this reason, we try to find measures of cognitive ability for both numeracy and literacy as far into adolescence as possible. At age 14 (sweep six), cohort members completed a Word Activity cognitive test. This specific assessment focused heavily on language acquisition and understanding of vocabulary, making it a useful control for literacy ability (Moulton et al., 2024). It was also measured during adolescence, where cognitive ability measures are considered highly stable, and just before the resit period after GCSE exams at age 16.

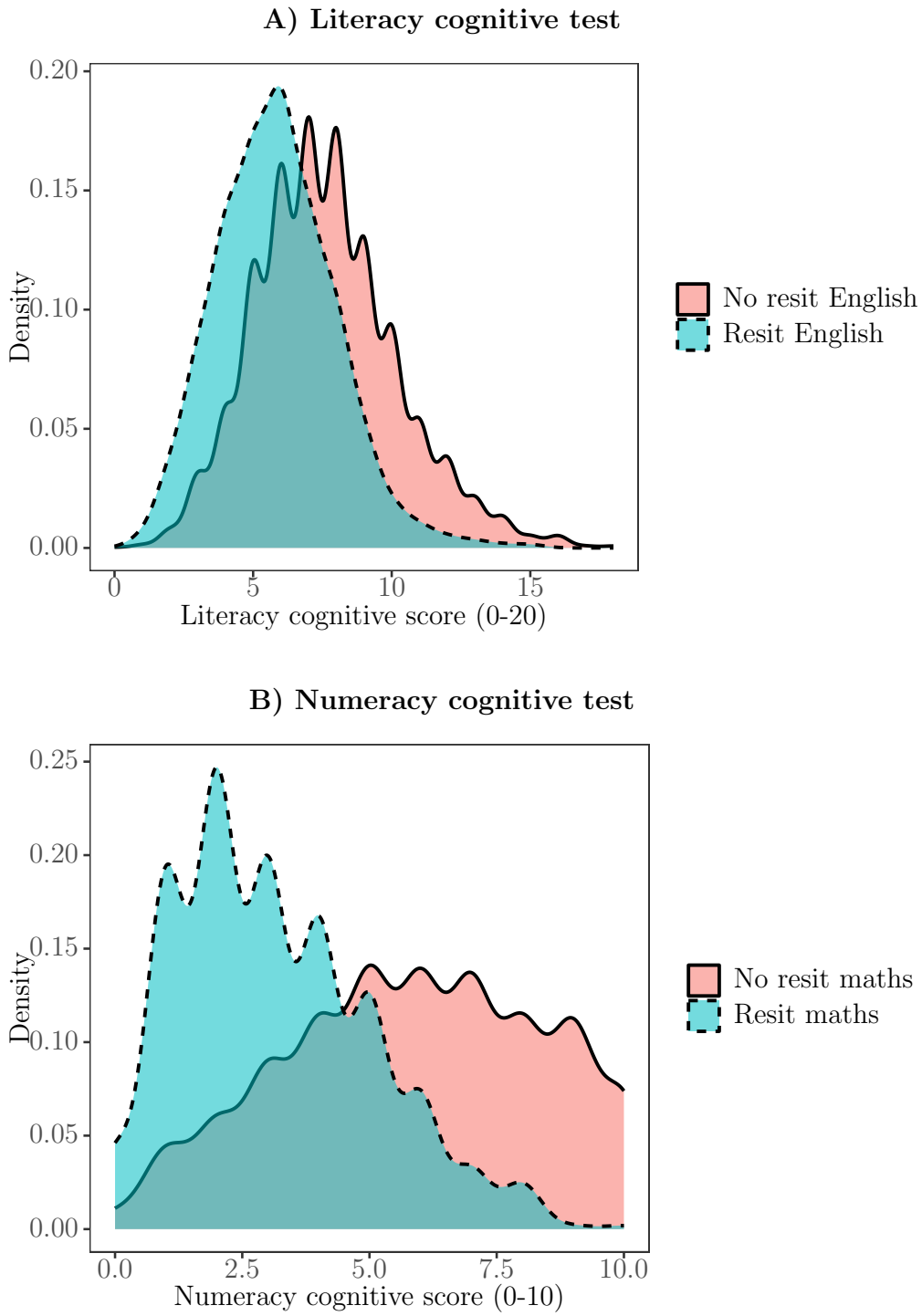
Our numeracy cognitive control comes from sweep seven which is collected at age 17. Cohort members completed a Number Analogies cognitive test that captures underlying quantitative reasoning and ability (Moulton et al., 2024). As such, it makes a useful control for numeracy ability. We use Sweep 7 because neither Sweep 6 (age 14) nor Sweep 5 (age 11) had clean cognitive ability tests that target numeracy. We would have to go back to Sweep 4 (age 7) for the next latest measure of numeracy ability.

However, one concern is that the Sweep 7 measure is taken during the resit period and could be influenced by the resits themselves. We are not overly concerned about this because research has shown these cognitive ability measures are highly stable once a child reaches adolescence (Breit et al., 2024; Gustavson et al., 2025). Moreover, if resits improve cognitive ability and ability is positively associated with wellbeing (i.e., Jokela, 2022), we would expect this to introduce a positive spurious association between resits and wellbeing. This means such confounding would under-estimate the relationships we observe in this paper. In addition, when we use age seven cognitive controls as a robustness check, the results remain unchanged.

Figure 4 shows the distribution of literacy and numeracy cognitive scores by whether a student is resitting in English or maths. As expected, there is a clear shift in the distribution

by resit status. Students who have to resit English score significantly worse on the literacy cognitive ability measure and students who resit maths score significantly worse on the numeracy ability measure. This is a reassuring sense check that our cognitive controls are capturing ability that is relevant to English and maths. Table 1 also illustrates this, showing it is those who fail English with lower literacy score and those who fail maths with lower numeracy scores. Failing English does not have a large association with numeracy ability and vice versa.

Figure 4: Distribution of literacy and numeracy cognitive scores by resit status



Notes: Figures show kernel density estimates of literacy and numeracy cognitive tests from the MCS for resit and non-resit students (in the relevant subject—English or maths). The numeracy measure is taken at age 17 (during the resit phase), whereas the literacy measure is taken at age 14, before students' high-stakes GCSE examinations at age 16.

4.4 Summary statistics

Table 1 presents descriptive statistics for the analysis sample by GCSE English and maths resit status. Overall, 18.4% of young people in the sample resat at least one GCSE, with 12.9% resitting English, 13.3% resitting maths, and 7.8% (373 individuals) resitting both subjects. Most respondents (81.6%) had passed both English and maths at age 16.

Students who resat GCSEs exhibited lower prior attainment than those who passed both subjects at age 16. Mean numeracy and literacy cognitive scores were substantially lower among resit students, particularly those who failed both subjects, while differences in mental wellbeing were comparatively modest. Resit students also came from less advantaged socioeconomic backgrounds, with average household income around £315 per week among those who failed both subjects compared with £453 among those who passed both GCSEs.

There are notable differences in the demographic composition of the resit groups. Males are overrepresented among students who failed both subjects (59.1%) and among those who passed maths but not English (67.0%), whereas females comprise the majority (72.0%) of students who passed English but not maths. The ethnic composition also varies across groups. White pupils account for around 72% of the full sample but a larger share of students who failed both subjects (75.5%) and those who passed maths but not English (79.0%). In contrast, Pakistani and Bangladeshi pupils are more heavily represented among those who passed English but not maths (16.7%) than in the full sample (10.8%).

Patterns in other GCSE attainment (fails in other subjects) are consistent with expectations. Close to 90% of students who failed both English and maths also failed at least two other GCSE subjects, compared with just 12.5% of those who passed both subjects. Conversely, around 71% of students who passed both English and maths had no other GCSE failures, compared with fewer than 2% of those who failed both subjects.

These statistics also highlight the importance of controlling for SEND. SEND status is measured using the age-14 parent report of whether the young person had been identified by their school or local authority as having special educational needs. Parents are then asked whether their child has an EHCP or is being assessed for one (EHCP category). We can see that SEND students are more likely to be resitting than non-SEND students, particularly for English (where 13.1% of the sample have SEND compared with 5.4% in the overall sample). As SEND is also correlated with wellbeing, it is important to account for this when looking at the relationship between retakes and wellbeing.

Table 1: Descriptive statistics by GCSE resit status

	Full sample (N = 4,776)	**Resit both** (N = 373)	**Resit neither** (N = 3,896)	**Resit maths** (N = 263)	**Resit English** (N = 244)
Cognitive and wellbeing measures					
Number Activity score	5.5 (2.6)	3.1 (2.0)	5.9 (2.5)	3.1 (1.8)	4.8 (2.4)
Word Activity score	7.4 (2.6)	5.5 (2.0)	7.7 (2.6)	6.5 (2.4)	6.3 (2.3)
Mental wellbeing	22.5 (4.1)	21.9 (4.2)	22.6 (4.0)	22.1 (4.6)	22.5 (3.9)
Prior wellbeing	33.0 (6.5)	32.1 (6.7)	33.3 (6.4)	31.3 (7.7)	32.3 (6.7)
Household income	434.0 (184.0)	315.6 (136.7)	453.2 (184.0)	349.4 (160.1)	387.2 (170.2)
Household size	4.5 (1.3)	4.6 (1.5)	4.4 (1.3)	4.5 (1.4)	4.5 (1.4)
Ethnicity (%)					
White	71.7	75.5	71.5	63.2	79.0
Mixed	6.0	7.2	5.6	8.4	8.5
Indian	4.4	0.9	4.9	3.4	2.7
Pakistani and Bangladeshi	10.8	10.5	10.8	16.7	5.8
Black or Black British	4.1	2.4	4.4	5.0	0.5
Other ethnic group	3.0	3.6	2.9	3.4	3.6
Sex (%)					
Male	47.0	59.1	45.9	28.0	67.0
Female	53.1	40.9	54.2	72.0	33.0
SEND (%)					
No known SEND	94.6	81.5	96.5	91.6	86.9
SEND (support)	3.2	9.1	2.2	4.6	9.0
EHCP	2.2	9.4	1.3	3.8	4.1
Prior GCSE failures (%)					
English resit	12.9	100.0	0.0	0.0	100.0
Maths resit	13.3	100.0	0.0	100.0	0.0
Any resit	18.4	100.0	0.0	100.0	100.0
No fails	59.7	1.4	70.9	17.6	15.2
One fail	16.5	10.3	16.6	24.1	16.5
Two or more fails	23.8	88.4	12.5	58.4	68.3

Notes: Continuous variables are means with standard deviations in parentheses; categorical variables are column percentages. "Other ethnic group" includes Chinese and other ethnic groups. Prior wellbeing and some covariates have missing observations. "Fails" refers to GCSE subjects other than English and mathematics.

5 Methods

We compare the wellbeing of resit students with non-resit students. A simple comparison across groups would be naive as the groups will differ in a range of other characteristics. We can see this in Table 1.

Fortunately, the MCS contains rich information on the students and their families which allow us to account for potential confounders. Specifically, we can control for **literacy and numeracy cognitive scores** taken very close to the resit period. This is extremely useful because we can compare students with similar underlying cognitive ability but one falls into the resit cohort and one doesn't. We also control for socioeconomic background (family size, income, housing tenure), sex, ethnicity and prior wellbeing.

The model we specify for individual i in household k is:

$$Y_{ik} = \alpha + \beta_1 Resit^{\mathbf{e}}_{ik} + \beta_2 Resit^{\mathbf{m}}_{ik} + \beta_3 Cog^{\mathbf{e}}_{ik} + \beta_4 Cog^{\mathbf{m}}_{ik} + \delta \mathbf{X}_{ik} + \omega \mathbf{Z}_k + \varepsilon_{ik} \tag{3}$$

where Y_{ik} is the wellbeing score for individual i in household k, $Resit^e_{ik}$ and $Resit^m_{ik}$ flag if the individual is resitting English or maths, Cog^e_{ik} records the literacy cognitive score achieved in sweep 6 of the MCS, Cog^m_{ik} records the maths cognitive score achieved in sweep 7 of the MCS, $\mathbf{X}_{ik}$ is a vector of individual characteristics (sex, ethnicity, month of survey, SEND status and prior wellbeing variables), $\mathbf{Z}_k$ is a vector of household characteristics (income, household size, housing tenure) and ε_{ik} is the idiosyncratic error term.

All models are estimated using the MCS survey design, incorporating the relevant Sweep 7 overall survey weights, primary sampling units and strata. These weights account for unequal selection probabilities arising from the original MCS sampling design and adjust for subsequent survey non-response. For the main England analysis, we use the single-country overall weight (`GOVWT1`). Our results are largely robust to using unweighted models.

5.1 Identification

Identifying the causal effect of resits on mental wellbeing relies on the assumption that, conditional on observed cognitive ability, prior wellbeing, demographic characteristics, and family background, assignment to the resit cohort is independent of potential wellbeing outcomes at age 17. Formally:

$$(Y_{ik}(1), Y_{ik}(0)) \perp Resit_{ik} \mid Cog_{ik}, X_{ik}, Z_k, \tag{4}$$

where $(Y_{ik}(1))$ and $(Y_{ik}(0))$ denote the potential wellbeing outcomes under resitting and not resitting, respectively. Under this assumption, any remaining differences in wellbeing

between otherwise similar students can be attributed to differences in resit status rather than to confounding factors. In other words, after conditioning on observed cognitive ability, demographic characteristics, prior wellbeing and family background, resit and non-resit students are assumed to be comparable in terms of their underlying determinants of wellbeing. Conditioning on prior wellbeing accounts for baseline differences in mental wellbeing between students, reducing the scope for reverse causality and omitted variable bias arising from persistent differences in underlying mental health.

The identifying assumption would be violated if there remain unobserved factors that jointly influence both GCSE resit participation and subsequent wellbeing, such as motivation, school support, or contemporaneous life events that are not captured by the observed covariates.[2] Our estimates should therefore be interpreted as causal under the maintained conditional-independence assumption rather than as arising from quasi-random assignment.

We provide an additional falsification test using Wales and Northern Ireland. Students in these countries could fail English or mathematics GCSEs but were not subject to the equivalent national compulsory resit requirement. This allows us to assess whether the England estimates may instead reflect the wellbeing consequences of examination failure itself. We discuss this falsification test in Section 5.4.

5.2 Theory-guided heterogeneity

Motivated by Section 3, we examine several dimensions of heterogeneity.

First, we interact the English and mathematics resit indicators with sex. These models test whether the subject-specific gender pattern anticipated by the framework is present in the data.

Second, we examine heterogeneity by distance from the pass threshold. We distinguish students who failed by one grade from those who failed by more than one grade. These estimates assess the competing hypotheses that wellbeing costs are larger where eventual success appears less attainable or, alternatively, where a narrowly missed pass is particularly salient.

Third, we interact resit status with SEND measured at age 14. The primary measure distinguishes students with no identified SEND from those receiving SEND support. In supplementary specifications, we separately identify students with an Education, Health and Care Plan or undergoing assessment for one. Because SEND encompasses heterogeneous needs and the relevant subgroups are small, these results are interpreted cautiously.

[2]However, if resitting means students end up at different types of post-16 institutions (Maris and Robinson, 2025) and this influences wellbeing, this would be a pathway through which resits influence wellbeing, not a confounding variable.

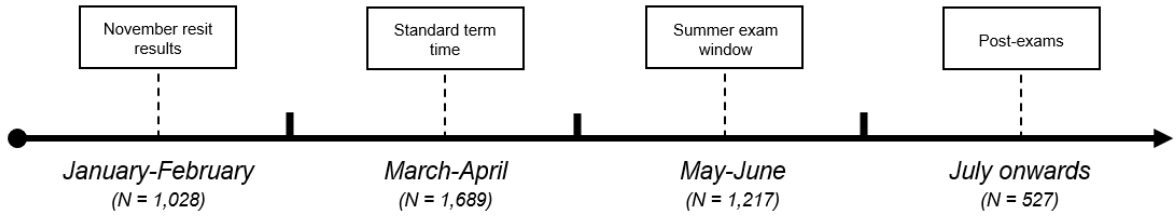
Finally, we interact resit status with the timing of the Wave 7 interview. Interview periods are grouped into January–February, March–April, May–June and July onwards (see Figure 5).

January–February is the period closest to when students originally received their results (mid–late August in the previous year). This is also a very salient period for students who entered a November resit immediately after starting their post–16 education (see Section 2 or Maris and Robinson (2025)). This is because these students receive their results in mid-January and most students are exposed to a second failure (the small number that pass likely will not be in our resit sample - see Section 4).

The March–April group is in between receiving results from November resits and the standard Summer exam window.

The May–June group captures the period closest to the summer examination series for resits and the post–July group captures all interviews after the standard Summer exam window.

Figure 5: Timeline for MCS age 17 sweep in England with respect to the resits cycle



These analyses do not directly identify the psychological mechanisms in the framework (see Section 3). Instead, they assess whether the pattern of heterogeneity is consistent with differences in expected success, subject-specific academic identity, additional needs and examination pressure.

5.3 Wellbeing dimensions and potential mechanisms

The Short Warwick–Edinburgh Mental Wellbeing Scale combines seven distinct dimensions of wellbeing. Examining these items separately allows us to assess which dimensions underpin any aggregate effects and whether the pattern is consistent with the mechanisms outlined in Section 3.3.

Our conceptual framework suggests that compulsory resitting may affect wellbeing through academic confidence, perceived competence, expectations about future success and examination-related stress. We therefore expect stronger relationships for items with a relatively direct connection to these channels, such as feeling useful and optimistic about the future. Other

dimensions, such as social connectedness and general coping ability, have a less direct theoretical relationship with resitting. The hypothesised links between resitting and each wellbeing item are summarised in Table 2.

Table 2: Potential pathways linking GCSE resits to dimensions of wellbeing

Wellbeing item	**Potential pathway**
Optimistic about the future	Repeated failure may reduce expectations about future success and progression (Avedissian and Alayan, 2021; Machin et al., 2020).
Feeling useful	Repeated failure and compulsory study may reinforce feelings of low competence and self-worth (Avedissian and Alayan, 2021; Crisp et al., 2023; Maris and Robinson, 2025).
Thinking clearly	Examination anxiety and academic pressure may affect concentration and perceived cognitive functioning (Fishstrom et al., 2022; Jerrim, 2023).
Feeling relaxed	Continued preparation for high-stakes examinations may increase stress and anxiety (Fishstrom et al., 2022; Jerrim, 2023).
Able to make up my own mind	Compulsory study and repeated failure may affect confidence, self-efficacy and perceived autonomy (Avedissian and Alayan, 2021).
Feeling close to other people	Resitting may indirectly affect social connectedness through disengagement, stigma or changes in students' educational experiences and peer interactions (Crisp et al., 2023; Maris and Robinson, 2025).
Dealing with problems well	Repeated failure and compulsory study may indirectly reduce perceived coping ability through lower confidence and self-efficacy (Avedissian and Alayan, 2021; Crisp et al., 2023).

Notes: The pathways are motivated by the conceptual framework in Section 3.3. Item-level estimates indicate which dimensions of wellbeing contribute to the aggregate association, but do not directly identify the underlying mechanisms.

5.4 Cross-country falsification test

We conclude the empirical analysis by replicating the main and heterogeneity specifications using cohort members in Wales and Northern Ireland (N = 1,401 in the preferred specification). These countries provide a falsification test because young people could fail English or mathematics GCSEs at age 16 without being subject to the same national compulsory resit

requirement as in England.[3]

We first estimate the average association between falling below the English or mathematics pass threshold and wellbeing. We then replicate the item-level, gender, distance-from-threshold, SEND and interview-timing analyses. This allows us to assess not only whether the average English association is absent outside England, but also whether the more specific pattern of heterogeneity observed in the English sample is reproduced.

If the English results primarily reflect the psychological consequences of examination failure, we would expect broadly similar average and subgroup patterns in Wales and Northern Ireland. By contrast, weaker or absent associations across both the main estimates and the heterogeneity analyses would be consistent with the English resit policy generating an additional wellbeing cost.

For the pooled falsification test, we use the multi-country overall weight which is recommended when analysing more than one country (`GOVWT2`). In constructing this sample, we identified a small number of respondents whose country of residence differed from the country in which they were originally sampled by the MCS. Some of these respondents retained relatively large weights associated with their original sampling stratum and, in the substantially smaller comparison-country sample, exerted disproportionate influence on the weighted estimates and substantially reduced the weight-based effective sample size. We therefore restrict the Wales and Northern Ireland analysis to respondents whose country corresponds to their original MCS sampling country, excluding 33 respondents (approximately 2% of the sample). For consistency, we apply the equivalent restriction to the England sample as a sensitivity check; doing so has no substantive effect on the England results.

6 Results

6.1 Average effects

Tables 3, 4 and 5 report the average effects of exposure to compulsory resits and mental wellbeing. The dependent variable is standardised, so the coefficients are expressed in standard-deviation units (SDs). Estimates are evaluated relative to students who do not do any resits. Table 3 uses a single indicator flagging if a respondent is required to do any English or mathematics resit. Table 4 interacts this with an indicator for having to do *both* English and maths resits. This provides evidence on whether there are additive effects of

[3]Some nevertheless chose, or were required by an individual institution or course, to retake to improve their scores. For example, a pass in English or mathematics may be an entry requirement for particular programmes or subsequent progression.

doing multiple resits. Table 5 provides a more detailed breakdown of the effects of resitting only English, only mathematics and both at the same time.

Using a single indicator for whether the respondent is subject to an English or mathematics resit requirement, our preferred specification estimates an average effect of 0.114 SDs on wellbeing. The estimate is statistically significant at the 5% level. It indicates that young people required to continue studying at least one of the two subjects report lower wellbeing than observationally similar young people who passed both subjects. This effect size is around one-third to one-half the average effect of school-based interventions explicitly designed to improve adolescent wellbeing, which recent meta-analyses estimate at approximately 0.24–0.32 SDs (Benavente et al., 2026; Tejada-Gallardo et al., 2020).

Table 3: Effects of any resit on wellbeing

	Base	Cognitive	Controls	Prior wellbeing
Resit	−0.053	−0.142***	−0.156***	−0.114**
	(0.093)	(0.053)	(0.054)	(0.050)
Observations	4776	4464	4391	4358
R^2	0.000	0.007	0.051	0.143
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Our results in Table 4 demonstrate that the negative wellbeing impacts arise from exposure to the compulsory resit requirement itself (extensive margin). There is no increasing wellbeing penalty for being required resit more than one subject. The interaction effect of resitting both is exactly zero in our preferred specification.

The results Table 5 largely support this interpretation. While the effect size on exposure to compulsory resits loses precision, we can see the effects for English (0.134 SDs), maths (0.095 SDs) and both (0.114 SDs) are relatively similar. While the coefficient on English is the only one that is statistically significant, all three coefficients are statistically indistinguishable

from one another and provide evidence that there is no additive effect from doing both English and maths resits.

Table 4: Number of resits on wellbeing

	Base	Cognitive	Controls	Prior wellbeing
Any resit	0.088	−0.159***	−0.178***	−0.114**
	(0.144)	(0.061)	(0.062)	(0.056)
Any resit × both	−0.304*	0.042	0.056	0.000
	(0.163)	(0.095)	(0.094)	(0.089)
Observations	4776	4464	4391	4358
R^2	0.005	0.007	0.052	0.143
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 5: Effects of resit group on wellbeing

	Base	Cognitive	Controls	Prior wellbeing
English resit	0.059	−0.155*	−0.218**	−0.134*
	(0.196)	(0.086)	(0.092)	(0.076)
Maths resits	0.114	−0.163**	−0.137*	−0.095
	(0.209)	(0.079)	(0.077)	(0.076)
Both resits	−0.216**	−0.118	−0.122	−0.114
	(0.087)	(0.083)	(0.083)	(0.081)
Observations	4776	4464	4391	4358
R^2	0.005	0.007	0.052	0.143
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

6.2 Falsification test - Wales and Northern Ireland

One of the challenges with identifying the impacts of compulsory resits on wellbeing is separating the impacts of being required to resit and the impacts of initial failure.

One of our strongest tests of our identification strategy is to replicate our analysis for young people in the MCS from Wales and Northern Ireland (N = 1,401 in preferred specification). This is a falsification test because students in Wales and Northern Ireland also take GCSE exams at age 16 (and English and maths are compulsory). However, unlike England, these countries did not have mandatory English and maths retakes. We do not include Scotland because they have a separate qualifications system and students do not sit GCSE exams at age 16.

We replicate all our models using the Welsh and Northern Ireland sample, including heterogeneity analyses, which we present and discuss across the remaining results sections. The results in Appendix C are clear – failing English and maths has a near-zero relationship

with wellbeing. The coefficient on failing at least one of English and maths in the Welsh and Northern Ireland sample is 0.028 SDs compared to the coefficient in the England sample of −0.114 SDs.

For our main results, we also conduct a formal test of whether the coefficient in England is different to Northern Ireland and Wales. We do this by running a pooled model, using the relevant cross-country MCS weights, and interact resitting with England and Wales/NI indicators.[4] We present the results below in Table 6. Failing either English or maths in England reduces wellbeing by 0.109 SDs. In Wales and NI, the effect is small, positive and statistically insignificant (0.034). Although the point estimates differ by 0.143 SDs, a formal test of equality does not reject that the coefficients are equal (p-value=0.17).

Table 6: Resit effects on wellbeing for Wales and Northern Ireland compared to England

	Base	Cognitive	Controls	Prior wellbeing
Resit × Wales/NI	0.038	0.003	0.025	0.034
	(0.085)	(0.084)	(0.096)	(0.098)
Resit × England	−0.054	−0.137***	−0.147***	−0.109**
	(0.094)	(0.052)	(0.053)	(0.049)
Observations	6293	5899	5804	5759
R^2	0.002	0.008	0.052	0.144
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, housing tenure, ethnicity, sex, SEND status, month of interview and country of residence. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Overall, these falsification tests provide support for our identification strategy. In a setting where students can fail the same exams but are not subject to resit requirements, we see no measurable impacts of failing on adolescent wellbeing. This supports the assertion

[4]To avoid interpreting more complex three-way or four-way interactions, we do not run these models for all the sub-group analyses.

that our results are capturing a resit effect rather than simply a more generic effect of failing a high-stakes examination (or resitting under voluntary conditions).

6.3 Gender heterogeneity

Table 7 examines whether the relationship between resit status and wellbeing differs by sex and subject. Due to strong gender differences in achievement and academic self-concept by subject (see Section 3), we run all our models using separate English and maths indicators.

Our estimates show a pronounced subject-specific pattern.

Among male respondents, English resit status is linked to significantly lower wellbeing (by 0.166 SDs) in the preferred specification. The corresponding English coefficient for female respondents is positive, small and statistically insignificant.

The pattern is reversed for mathematics. Mathematics resit status is linked with lower wellbeing for female respondents (by 0.142 SDs). In contrast, there is a relatively small positive and statistically insignificant estimate among male respondents.

These findings are consistent with subject-specific differences in academic self-concept and examination anxiety discussed in Section 3. They do not, however, directly establish those psychological mechanisms. The corresponding Wales and Northern Ireland estimates (falsification tests) do not reproduce this subject-specific gender pattern (Table C5).

Table 7: Gender heterogeneity in resit effects on overall wellbeing

	Base	Cognitive	Controls	Prior wellbeing
Resit maths × male	0.091	0.077	0.112	0.065
	(0.106)	(0.105)	(0.110)	(0.097)
Resit English × male	−0.268***	−0.297***	−0.245**	−0.166*
	(0.100)	(0.097)	(0.103)	(0.084)
Resit maths × female	−0.172**	−0.188**	−0.174**	−0.142*
	(0.079)	(0.081)	(0.083)	(0.080)
Resit English × female	−0.037	−0.061	0.010	−0.004
	(0.101)	(0.102)	(0.102)	(0.099)
Observations	4464	4464	4391	4358
R^2	0.040	0.042	0.053	0.143
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. The table reports interactions between GCSE resit indicators and sex. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

6.4 Individual wellbeing items

Table 8 examines the seven components of the wellbeing scale separately. All specifications include cognitive scores, demographic and socioeconomic characteristics, interview timing and prior wellbeing. Table 9 provides the same results with gender interactions.

Across the two tables, the negative effects are strongest for thinking clearly, making your mind up and feeling useful. The gender patterns we identified with our overall outcome variable in Table 7 are replicated here. Females who resit maths report feeling significantly less useful and able to think clearly than other female students who are not resitting. Males who resit English report feeling significantly less useful, able to think clearly, make their own mind up and report feeling less close to others. We find no statistically significant rela-

tionships between any of the items and resitting maths for male students and no significant associations with resitting English for female students.

The corresponding Wales and Northern Ireland estimates (falsification tests) show no statistically significant effects on any items for English or mathematics (Table C4).

Table 8: Resit effects on individual wellbeing items

	Optimism	Useful	Thinking clearly	Relaxed	Make mind up	Close to others	Problems
Resit maths	−0.077	−0.094	−0.166**	0.000	0.005	0.071	−0.060
	(0.083)	(0.078)	(0.072)	(0.057)	(0.073)	(0.086)	(0.124)
Resit English	−0.089	−0.106	−0.046	0.045	−0.143*	−0.107	0.054
	(0.089)	(0.070)	(0.068)	(0.063)	(0.079)	(0.069)	(0.093)
Observations	4358	4358	4358	4358	4358	4358	4358
R^2	0.057	0.106	0.121	0.139	0.056	0.072	0.096
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Outcome mean	3.43	3.33	3.50	3.32	3.97	3.71	3.37

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variables are the individual items from the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 9: Resit effects on individual wellbeing items - gender interactions

	Optimism	Useful	Thinking clearly	Relaxed	Make mind up	Close to others	Problems
Resit maths - Male	−0.028	0.085	0.043	0.135	0.019	0.225	−0.154
	(0.105)	(0.118)	(0.091)	(0.083)	(0.107)	(0.151)	(0.230)
Resit maths - Female	−0.112	−0.236**	−0.324***	−0.099	−0.002	−0.049	0.013
	(0.100)	(0.092)	(0.105)	(0.086)	(0.095)	(0.074)	(0.098)
Resit English - Male	−0.135	−0.197**	−0.126*	−0.009	−0.150	−0.191*	0.088
	(0.100)	(0.092)	(0.066)	(0.071)	(0.091)	(0.108)	(0.154)
Resit English - Female	−0.050	−0.011	0.016	0.081	−0.144	−0.025	0.026
	(0.148)	(0.112)	(0.117)	(0.098)	(0.132)	(0.073)	(0.115)
Observations	4358	4358	4358	4358	4358	4358	4358
R^2	0.057	0.108	0.124	0.140	0.056	0.073	0.096
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Outcome mean	3.43	3.33	3.50	3.32	3.97	3.71	3.37

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variables are the individual items from the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

6.5 Distance from the pass threshold

The framework generates two competing predictions about distance from the pass threshold. Students further below the threshold may experience greater discouragement because eventual success appears less attainable. Conversely, students who narrowly fail may experience a particularly salient near miss. Given distance from the pass threshold is subject-specific and our subject-specific results are strongly gendered, we present our results by gender in Table 10.

Interestingly, we find support for the discouragement hypothesis for males and support for the near miss hypothesis for females.

Female students who were resitting maths and missed out by one grade have 0.19 SDs lower wellbeing than non-resitting females. In contrast, the effect size is much lower (0.039 SDs) and not significant for females who were far from the threshold. A narrow miss may be especially salient for girls because girls tend to report higher mathematics anxiety than boys even when performance is similar (Devine et al., 2012).

For males resitting English, the pattern is more consistent with the discouragement hypothesis. Although the relationship is negative among students who narrowly missed the threshold, it is significantly larger for those further below it. Male students further from the pass threshold have wellbeing scores 0.308 SDs lower than comparable non-resitting males, almost three times the effect size for those only one grade below the passing threshold. One possible interpretation is that for males who are already relatively far from a pass in English, the prospect of eventual success appears less attainable. This may be having a discouragement effect for males, reinforcing pre-existing low English self-concept (Nagy et al., 2010; Sullivan, 2009).

The corresponding Wales and Northern Ireland estimates (falsification tests) show no statistically significant effects at any distance from the passing threshold (Table C7).

Table 10: Resit effects on wellbeing for close vs not close fails by gender

	Base	Cognitive	Controls	Prior wellbeing
Mathematics				
Maths: close × male	0.105	0.101	0.132	0.115
	(0.122)	(0.120)	(0.127)	(0.106)
Maths: close × female	−0.208**	−0.222**	−0.205**	−0.186**
	(0.095)	(0.095)	(0.095)	(0.086)
Maths: not close × male	0.143	0.119	0.153	0.055
	(0.145)	(0.144)	(0.147)	(0.136)
Maths: not close × female	−0.091	−0.111	−0.109	−0.039
	(0.147)	(0.151)	(0.147)	(0.129)
English				
English: close × male	−0.214**	−0.240**	−0.192*	−0.112
	(0.098)	(0.095)	(0.098)	(0.090)
English: close × female	−0.061	−0.084	−0.020	−0.031
	(0.113)	(0.113)	(0.110)	(0.095)
English: not close × male	−0.425**	−0.459***	−0.397**	−0.308**
	(0.173)	(0.168)	(0.179)	(0.133)
English: not close × female	−0.043	−0.068	0.034	−0.021
	(0.226)	(0.226)	(0.227)	(0.247)
Observations	4464	4464	4391	4358
R^2	0.041	0.043	0.053	0.145
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, sex, SEND status and month of interview. Among English resit students, 392 respondents were one grade below the pass threshold and 167 were more than one grade below it. The corresponding numbers for mathematics are 341 and 233. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

6.6 SEND

Table 11 examines whether the effects of exposure to compulsory resits on wellbeing differs according to SEND status, measured at age 14. It is worth acknowledging that these interaction terms are based on a relatively small number of students (around 95% of the overall sample do not have SEND - see Table 1). Therefore, we have limited power to detect effects among the sub-groups with SEND.

SEND is broken down into students receiving a basic level of support (SEND support) and those with more substantial needs who have an Education, Health and Care Plan or were being assessed for one. Table 11 shows no statistically significant differences by SEND group. In the specifications without full controls, there is a sizeable negative coefficient on resitting English for those with an EHCP. However, these become statistically insignificant and imprecisely estimated in the final specification. There are also no statistically significant effects in the corresponding Wales and Northern Ireland models (falsification tests).

Table 11: Resit effects on wellbeing by SEND

	Base	Cognitive	Controls	Prior wellbeing
Resit maths × no SEND	−0.053	−0.121*	−0.082	−0.077
	(0.174)	(0.066)	(0.072)	(0.061)
Resit English × no SEND	−0.085	−0.049	−0.119	−0.087
	(0.171)	(0.079)	(0.080)	(0.070)
Resit maths × SEND support	0.301	0.316*	0.294	0.124
	(0.182)	(0.188)	(0.190)	(0.209)
Resit English × SEND support	−0.082	−0.088	−0.084	−0.045
	(0.178)	(0.183)	(0.186)	(0.185)
Resit maths × EHCP	0.147	0.162	0.163	0.249
	(0.194)	(0.199)	(0.214)	(0.260)
Resit English × EHCP	−0.338*	−0.352*	−0.347	−0.254
	(0.199)	(0.202)	(0.221)	(0.234)
Observations	4776	4464	4391	4358
R^2	0.006	0.010	0.052	0.143
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, sex, SEND status and month of interview. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

6.7 Timing within the resit cycle

Table 12 examines whether the effects of exposure to compulsory resits varies across the academic year. Effects by subject are presented in the Appendix (Table D1). The estimates reveal some limited variation by both subject and interview period.

Using the combined resit indicator, the negative effects are statistically significant among respondents interviewed in January or February. Resit status is linked to 0.168 SDs lower wellbeing in this period. This period coincides with the release of results from the November

examination series and is the closest period to the original results release in August of the previous year. The estimates for May–June, coinciding with the main resit exam period, are similar in size but not statistically significant. Estimates from July onwards (post-resit exams) and March–April (months preceding exams) are smaller and statistically insignificant.

The subject-specific models help clarify these patterns further in the Appendix (Table D1). Mathematics resit status is linked to lower wellbeing in January–February whereas English resit status is not. This may be because November resit entries are more common in maths than English (Maris and Robinson, 2025), so more maths resit students have received news about an additional failure. These results are consistent with the possibility that receiving another unsuccessful result contributes to lower wellbeing.

Our falsification tests using the Wales and Northern Ireland sample are of particular interest here. If we suspect our results may also be capturing an exam failure effect rather than a resitting effect, we would expect to see negative effects closer to the initial GCSE exams themselves (the Jan/Feb group), which dissipate over time as the failure becomes less salient. We do not observe this pattern in Table C8. None of the coefficients are significant and the Jan/Feb month actually has the most positive coefficient (not the most negative as we would expect).

Table 12: Resit effects on wellbeing by month

	Base	Cognitive	Controls	Prior wellbeing
Any resit × Jan/Feb	−0.024	−0.196*	−0.209**	−0.168*
	(0.239)	(0.105)	(0.101)	(0.091)
Any resit × March/Apr	−0.147*	−0.110	−0.123	−0.110
	(0.077)	(0.076)	(0.078)	(0.070)
Any resit × May/June	−0.071	−0.164	−0.161	−0.147
	(0.185)	(0.115)	(0.107)	(0.100)
Any resit × July on	0.036	−0.125	−0.157	0.002
	(0.260)	(0.134)	(0.150)	(0.121)
Observations	4776	4464	4391	4358
R^2	0.006	0.008	0.052	0.143
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, sex, SEND status and month of interview. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

The item-level timing estimates in Appendix Table G1 provide further context. For mathematics, the January–February effect is concentrated in optimism (−0.555 SDs) and feeling useful (−0.233 SDs). For English, the May–June estimates are similarly concentrated in optimism and usefulness. From July onwards, English resit status is positively associated with thinking clearly, feeling relaxed and dealing with problems well. This configuration is consistent with acute examination- and results-related changes rather than a constant wellbeing difference throughout the year.

These interpretations remain tentative. Interview timing is not randomly assigned, not all mathematics resit students entered the November series, and the July-onwards category includes some interviews conducted in a subsequent post-16 year. Successful resitters may also cease to be classified as resit students. The timing estimates should consequently be treated as mechanism-consistent evidence rather than causal effects of particular stages of

the resit cycle.

7 Robustness checks

7.1 Tightening the comparison group

One potential concern with our main specification is that our comparison group of non-resitting students includes students who achieved substantially higher GCSE grades in those subjects. Although our models control flexibly for cognitive ability and a rich set of individual and family characteristics, comparisons with much higher-attaining students may be more prone to underlying differences in unobservables, which could present additional confounding risks.

We therefore repeat our main analysis using a more restricted comparison group. We retain all students subject to an English and/or mathematics resit requirement, but restrict non-resit students to those who achieved a grade 4 (or grade C) in English and/or mathematics. These are students who met the passing threshold by a single grade, making them substantially closer in attainment to the resit population than the full group of students who passed both subjects. This retains 32% of our sample of non-resitting students.

We find that our results become stronger when using this more comparable group of lower-attaining passers. In the preferred specification, the estimated effect of being subject to at least one resit requirement increases in magnitude from approximately 0.11 SDs in the full sample to approximately 0.18 SDs in the restricted comparison sample (Table 13).

The same pattern is evident when resit status is separated into English-only, mathematics-only and dual-subject resits. Each estimate becomes more negative when higher-attaining comparison students are excluded and all estimates become statistically significant (Table 13). The full modelling results can be found in Appendix H in Tables H1 and H2.

These results provide additional reassurance that the main findings are not driven by comparisons between resit students and substantially higher-attaining students. If anything, restricting the analysis to students closer to the GCSE pass threshold strengthens the estimated wellbeing penalty associated with compulsory resitting. The restricted analysis nevertheless uses a different, more academically marginal comparison population and should therefore be interpreted as a complementary estimate rather than a direct replacement for the full-sample result.

(Table H1).

Table 13: Resit effects using the full and restricted comparison groups

	Full comparison group	Restricted comparison group
Panel A: Any resit		
Any English or mathematics resit	**−0.114****	**−0.176*****
	(0.050)	(0.059)
Panel B: Resit group		
English only	**−0.134***	**−0.188****
	(0.076)	(0.085)
Mathematics only	-0.095	**−0.147***
	(0.076)	(0.081)
Both English and mathematics	-0.114	**−0.192****
	(0.081)	(0.086)
Observations	4,358	1,888

Notes: Survey-weighted estimates from the preferred fully controlled specification are reported. Survey-design-based standard errors are in parentheses. The full comparison group includes all respondents who passed both GCSE English and mathematics. The restricted comparison group retains all resit students but restricts non-resit respondents to those who achieved a grade 4 (or grade C) in English and/or mathematics. All models control for cognitive ability, household income, household size, housing tenure, ethnicity, SEND status, sex, month of interview and prior wellbeing. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

7.2 Broader academic failure

Our analysis of the Wales and Northern Ireland sample provide reassurance that we are identifying something meaningful in relation to links between England's resit policy settings and adolescent wellbeing. However, there still may be other underlying cross-country differences in education systems and institutions that could mean failing an exam in England has a negative wellbeing effect but failing an exam in Wales and Northern Ireland does not. While this seems unlikely, it remains a concern and limitation of our approach.

In an ideal world, we would be able to separate the effects of failing any GCSE exam with the effects of failing English or maths and being required to resit in the England sample only. We attempted to model this by including a categorical variable mapping how many additional GCSE failures a student reports (zero, one or two or more). Unfortunately, however, we cannot separately identify these effects. Because English and maths are core skills that feed into many other GCSEs, it is very rare to only fail English or only fail maths or only fail one or both of the two (Table 14).

There is only a very small and atypical group of resit students that only fail English and/or maths. Only 37 students failed English alone, 46 failed mathematics alone, and five failed both subjects without failing any other GCSEs. This means 90% of resit students

in our sample also failed at least one other subject (that they reported in the MCS – the number could be much higher if under-reporting is concentrated among GCSEs with lower grades).

Table 14: Composition of the GCSE resit group

Group	**N**	**Resit students (%)**	**All students (%)**
Resit English only	37	4.2	0.8
Resit maths only	46	5.3	1.0
Resit English and maths only	5	0.6	0.1
Resit English and other* subjects	206	23.6	4.3
Resit maths and other* subjects	216	24.7	4.5
Resit English, maths and other* subjects	364	41.6	7.6
Total	874	100.0	18.3

Notes: *"Other subjects" refers to GCSE subjects other than English and mathematics. Percentages of resit students are calculated among students resitting English and/or mathematics with non-missing information on other GCSE failures. Percentages of all students are calculated relative to the full sample ($N = 4,776$). Percentages may not sum exactly due to rounding.

For transparency, we present the results of our models with controls for additional exam failures in Appendix F. However, the estimand has changed dramatically and now represents the effects of not passing for those who reported **no other failures**. The effects for this atypical group tend to be in the same direction but much smaller in magnitude, less precise and often become statistically insignificant. However, some of the more specific item-level and interaction results remain significant. We remain focused on our main findings because of how uncommon failing *only* English or maths is, which limits the usefulness of this exercise.

7.3 Additional robustness checks

We conduct a further set of robustness checks. First, we replace the contemporaneous age-17 numeracy assessment with the earlier numeracy measure collected at age seven. This avoids conditioning on a measure that could in principle have been affected by post-16 study. The qualitative pattern of the results is unchanged, although the earlier assessment is a less proximate measure of adolescent numeracy.

Second, we allow English and mathematics resit status to interact. There is no evidence that being subject to both requirements has a non-additive association with wellbeing beyond the sum of the two subject-specific associations. Full estimates are reported in Appendix E.

Third, we run all of our models without accounting for MCS design-based weights. This is to address the concern that the weights themselves may be driving some of our results. We find our results are substantively stable to running weighted or unweighted models.

Fourth, we run models where we exclude respondents who were not initially in the England MCS sample (but moved to England by Sweep 7). We do this because the MCS survey weights are based on the country where the respondent's family was initially recruited. This results in the exclusion of N = 60 respondents (1.4% of the sample) and our results remain largely unchanged.

Finally, we exclude respondents in England who were surveyed after August of 2018 and may have thus already received results from their summer resits. These respondents would also be in their second-year of post-16 rather than their first (see Figure 1). This results in the exclusion of N = 148 respondents (3% of the sample) and our results remain stable.

8 Conclusion

Education policies frequently require low-attaining students to undertake additional study in order to reach minimum skill standards. Although the potential human-capital benefits of these requirements are clear, considerably less attention has been paid to their potential non-pecuniary costs.

Using England's compulsory GCSE resit requirement, we estimate that exposure to compulsory continued study lowers adolescent wellbeing by approximately 0.11 standard deviations (SDs). The causal interpretation of this estimate rests on conditional independence after accounting for prior wellbeing, cognitive ability and a rich set of observed characteristics. A cross-country falsification exercise provides additional evidence against a simple explanation based solely on the consequences of examination failure. We show that the average estimate masks substantial heterogeneity.

Mathematics resit status is more negatively associated with wellbeing among girls, while its largest timing-specific estimate occurs in January and February, coinciding with the November resit-results period. The negative association for girls is also concentrated among those who narrowly missed the mathematics pass threshold. This is consistent with a potential near-miss penalty: narrowly missing an attainable outcome may make the counterfactual of having passed particularly salient, increasing disappointment (Roese, 1997). This may be especially relevant in mathematics, given evidence that girls report lower mathematics self-concept and greater mathematics anxiety on average (Devine et al., 2012; Nagy et al., 2010; Sullivan, 2009). There are no detectable effects of resitting English for girls.

The pattern for boys resitting English is notably different. Here, we find a negative association between resitting and wellbeing that is concentrated among students further below the pass threshold. This is more consistent with a discouragement mechanism: where the perceived probability of eventually passing is low, the benefits of continued compulsory

study may appear less attainable relative to its psychological and effort costs. Qualitative evidence from the resit context similarly highlights low confidence, feelings of failure and disengagement among some resit students (Crisp et al., 2023; Maris and Robinson, 2025). However, interestingly, we only see these discouragement-style effects for boys resitting English. Taken together, these results suggest that both narrowly missing a pass and being substantially below the threshold may be consequential for wellbeing, but potentially for different students and through different processes. Further research should investigate these mechanisms directly.

The individual wellbeing items provide further evidence about the form of the relationship. Associations are concentrated in feeling useful and thinking clearly. By contrast, feeling close to other people and dealing with problems well, two of the outcomes with weaker theoretical connections to resitting, show little association.

Our cross-country falsification tests provide evidence that we are identifying effects linked to resits rather than exam failure itself. None of the average or subgroup patterns observed in England are reproduced in Wales and Northern Ireland where students also take GCSEs but face no equivalent national resit requirement.

Our paper has several limitations. Resit status is inferred from self-reported qualification records and may be measured with error. The identification strategy is based on conditional independence and cannot eliminate unobserved differences in motivation, attendance, school support or contemporaneous life events. Our falsification tests provide support for our identification but residual confounders may still remain. The age-17 numeracy assessment is collected during the resit period and may itself be affected by post-16 study. The timing estimates are additionally vulnerable to selective reclassification when successful resitters cease to appear in the resit group. Future research could focus on developing longitudinal data at a higher frequency that would enable researchers to follow students before results, after initial failure and through successive resit attempts. This would enable further investigation of the mechanisms underlying our headline results.

The findings do not imply that post-16 support for English and mathematics should be withdrawn. Securing these skills remains an important educational objective. They do suggest that policy evaluation should consider wellbeing alongside attainment and progression. Support may be particularly important around examination and results periods, both for students who are substantially below the pass threshold and those who remain disappointed after narrowly missing a pass. More differentiated teaching, credible alternative qualification routes, appropriate assessment adjustments and pastoral support could potentially preserve the policy's educational aims while reducing its costs to confidence and wellbeing.

More broadly, our findings highlight an important trade-off in policies that compel ad-

ditional human-capital investment after academic failure. Further study may improve skills and subsequent educational and labour-market opportunities, but the way that remediation is organised can also impose meaningful costs while that investment takes place. Evaluations of compulsory remedial education should therefore consider student wellbeing alongside attainment and progression.

References

Angrist, J. D. and Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, 106(4):979–1014.

AOC (2023). Pressure on colleges mounts as booming number of students resit GCSE maths and English.

Association of Colleges (2022). AoC English and maths survey January 2022. Technical report.

Avedissian, T. and Alayan, N. (2021). Adolescent well-being: A concept analysis. *International Journal of Mental Health Nursing*, 30(2):357–367.

Avendano, M., de Coulon, A., and Nafilyan, V. (2020). Does longer compulsory schooling affect mental health? Evidence from a British reform. *Journal of Public Economics*, 183:104137.

Beck, K. C., Røhr, H. L., Reme, B.-A., and Flatø, M. (2024). Distressing testing: A propensity score analysis of high-stakes exam failure and mental health. *Child Development*, 95(1):242–260. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cdev.13985.

Benavente, M., Rodríguez-Rivas, M. E., Valdebenito, S., Chuecas, M. J., and Alfaro, J. (2026). Effectiveness of School-Based Interventions on Adolescents' Subjective Well-Being: A Systematic Review and Meta-Analysis of Recent Evidence. *Child Indicators Research*.

Bizopoulou, A., Megalokonomou, R., and Simion, (2024). Do second chances pay off? Evidence from a natural experiment with low-achieving students. *Journal of Public Economics*, 239:105214.

Breit, M., Scherrer, V., Tucker-Drob, E. M., and Preckel, F. (2024). The stability of cognitive abilities: A meta-analytic review of longitudinal studies. *Psychological Bulletin*, 150(4):399–439.

Busby, E. (2024). School leaders condemn 'remorseless treadmill' of GCSE English and maths resits. *The Independent*. Section: News.

Camden, B. (2023). GCSE resits 2023: Maths and English pass rates down again.

Crisp, B., Hallgarten, J., Joshua, V., Morris, R., Perry, T., and Wardle, L. (2023). Post-16 GCSE Resit Practice Review. Technical report, EEF.

Denny, L. (2026). English and maths resits keep sucking life out of students. *FE Week*.

Devereux, P. J. and Hart, R. A. (2010). Forced to be Rich? Returns to Compulsory Schooling in Britain. *The Economic Journal*, 120(549):1345–1364.

Devine, A., Fawcett, K., Szűcs, D., and Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and brain functions: BBF*, 8:33.

Education and Skills Funding Agency and Department for Education (2025). 16 to 19 funding: maths and English condition of funding.

Finning, K., Ford, T., and Moore, D. A., editors (2022). *Mental Health and Attendance at School*. Cambridge University Press, Cambridge.

Fishstrom, S., Wang, H.-H., Bhat, B. H., Daniel, J., Dille, J., Capin, P., and Vaughn, S. (2022). A meta-analysis of the effects of academic interventions on academic achievement and academic anxiety outcomes in elementary school children. *Journal of school psychology*, 92:265–284.

Goodman, J., Gurantz, O., and Smith, J. (2020). Take Two! SAT Retaking and College Enrollment Gaps. *American Economic Journal: Economic Policy*, 12(2):115–158.

Gustavson, D. E., Borriello, G. A., Karhadkar, M. A., Rhee, S. H., Corley, R. P., Rhea, S.-A., DiLalla, L. F., Wadsworth, S. J., Friedman, N. P., and Reynolds, C. A. (2025). Stability of general cognitive ability from infancy to adulthood: A combined twin and genomic investigation. *Proceedings of the National Academy of Sciences*, 122(21):e2426531122.

Hatton, C., Emerson, E., Robertson, J., and Baines, S. (2018). The mental health of adolescents with and without mild/moderate intellectual disabilities in England: Secondary analysis of a longitudinal cohort study. *Journal of Applied Research in Intellectual Disabilities*, 31(5):768–777. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jar.12428.

Högberg, B. and Horn, D. (2022). National High-Stakes Testing, Gender, and School Stress in Europe: A Difference-in-Differences Analysis. *European Sociological Review*, 38(6):975–987.

Jacob, B. A. and Lefgren, L. (2004). Remedial Education and Student Achievement: A Regression-Discontinuity Analysis. *The Review of Economics and Statistics*, 86(1):226–244.

Jacob, B. A. and Lefgren, L. (2009). The Effect of Grade Retention on High School Completion. *American Economic Journal: Applied Economics*, 1(3):33–58.

Jerrim, J. (2023). Test anxiety: Is it associated with performance in high-stakes examinations? *Oxford Review of Education*, 49(3):321–341. _eprint: https://doi.org/10.1080/03054985.2022.2079616.

Jokela, M. (2022). Why is cognitive ability associated with psychological distress and well-being? Exploring psychological, biological, and social mechanisms. *Personality and Individual Differences*, 192:111592.

Joshi, H. and Fitzsimons, E. (2016). The Millennium Cohort Study: the making of a multi-purpose resource for social science and policy. *Longitudinal and Life Course Studies*, 7(4):409–430.

Krishna, K., Lychagin, S., and Frisancho, V. (2018). Retaking in High Stakes Exams:

Is Less More? *International Economic Review*, 59(2):449–477. Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University].

Machin, S., McNally, S., and Ruiz-Valenzuela, J. (2020). Entry through the narrow door: The costs of just failing high stakes exams. *Journal of Public Economics*, 190:104224.

Maris, R. (2024). Time for a resit reset? EPI Blog. https://epi.org.uk/publications-and-research/blog-time-for-a-resit-reset/.

Maris, R., Khandekar, S., and Robinson, D. (2024). A quantitative analysis of T level access and progression. Technical report, Education Policy Institute.

Maris, R. and Robinson, D. (2025). English and Maths Resits - Drivers of Success. Technical report, Education Policy Institute.

Monk, F. (2025). It's time for a GCSE resits rethink says Pearson.

Moulton, V., McElroy, E., Richards, M., Fitzsimons, E., Northstone, K., Conti, G., Ploubidis, G., Sullivan, A., Başтuğ, S., Sykes, K., Campbell, C., and O'Neill, D. (2024). A guide to the cognitive measures in five CLOSER studies. Technical report, CLOSER, London, UK.

Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., and Baumert, J. (2010). The Development of Students' Mathematics Self-Concept in Relation to Gender: Different Countries, Different Trajectories? *Journal of Research on Adolescence*, 20(2):482–506. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532-7795.2010.00644.x.

OECD (2013). England & Northern Ireland (UK) – Country Note –Survey of Adult Skills first results. Technical report.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, 121(1):133–148.

Stopforth, S., Connelly, R., and Gayle, V. (2025). Reporting and measuring English school qualifications: a case study of General Certificate of Secondary Education results in survey and linked administrative data in the UK Millennium Cohort Study. *Cambridge Journal of Education*, 55(6):871–890. _eprint: https://doi.org/10.1080/0305764X.2025.2552886.

Sullivan, A. (2009). Academic self-concept, gender and single-sex schooling. *British Educational Research Journal*, 35(2):259–288. _eprint: https://doi.org/10.1080/01411920802042960.

Tafreschi, D. and Thiemann, P. (2016). Doing it twice, getting it right? The effects of grade retention and course repetition in higher education. *Economics of Education Review*, 55:198–219.

Tejada-Gallardo, C., Blasco-Belled, A., Torrelles-Nadal, C., and Alsinet, C. (2020). Effects of School-based Multicomponent Positive Psychology Interventions on Well-being and Distress in Adolescents: A Systematic Review and Meta-analysis. *Journal of Youth*

and Adolescence, 49(10):1943–1960.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., and Stewart-Brown, S. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes*, 5.

The Royal Society (2021). The Qualifications Assessment Framework and GCSE Mathematics resits: A working paper. Technical report.

Tuckett, S. (2025). Post 16 study programmes – understanding student choices and aspirations.

von der Embse, N., Jester, D., Roy, D., and Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, 227:483–493.

Appendix A. Consistency of the self-reported GCSE data

We are using self-reported grades from the MCS. These may differ from the grades reported in the National Pupil Database – England's administrative dataset for students attending state–funded schools. Stopforth et al. (2025) showed there was some inconsistency between self-reported grades in the MCS and grades recorded in the administrative data. There was some measurement error in both the positive and negative direction (and missingness), but their example revealed this had little impact on their empirical results (Stopforth et al., 2025). Because students do somewhere between eight and ten GCSEs on average, there is substantial room for error when reporting entries and grades for all GCSEs. However, English and mathematics are two of the compulsory GCSEs and are more likely to be reported without error (Stopforth et al., 2025).

We consider the consistency in self-reported grades for English and maths, the two most relevant GCSEs to this paper. We use data presented by Stopforth et al. (2025). Table A1 reports the results from this exercise. Some students convert their numeric grades to an alphabetical one when reporting in the MCS. Because many numeric grades do not have an exact alphabetical equivalent (six out of the nine do not have exact equivalents), some error in converting is to be expected (and is not necessarily an error as the grade reported may be the closest to an exact equivalent).

We can see that of those with data in both the MCS and NPD, 85% match exactly for English language and 89% match for mathematics. If we consider the issues with equivalence when converting grades from numeric to alphabetical values, the match rate could be closer to 89% for English and 92% for maths. The most significant other error arises from students reporting a higher grade by one unit (6% for English and 5% for maths). Misreporting by larger amounts is uncommon (3%).

Overall, these error rates are low and we are re-assured by the empirical examples in Stopforth et al. (2025) showing that these measurement errors make little difference to the example model estimates. Moreover, our primary interest is around the cutoff for doing resits (grade 4). The only measurement error that would affect this variable is measurement error right around this threshold, meaning the estimates in Table A1 are conservative. The salience of this threshold also makes it plausible that we would see less measurement error around this cutoff.

Table A1: Agreement between MCS and NPD attainment records

	English language		Mathematics	
	Number	Percent	Number	Percent
Exact match[a]	4,021	85%	4,287	89%
Higher alphabetical grade	27	1%	16	0%
Lower alphabetical grade	155	3%	124	3%
Higher grade	273	6%	217	5%
Lower grade	97	2%	25	1%
More than one grade difference	158	3%	127	3%
Match including adjacent alphabetical grades[b]	4,203	89%	4,427	92%
Total	4,731	100%	4,796	100%

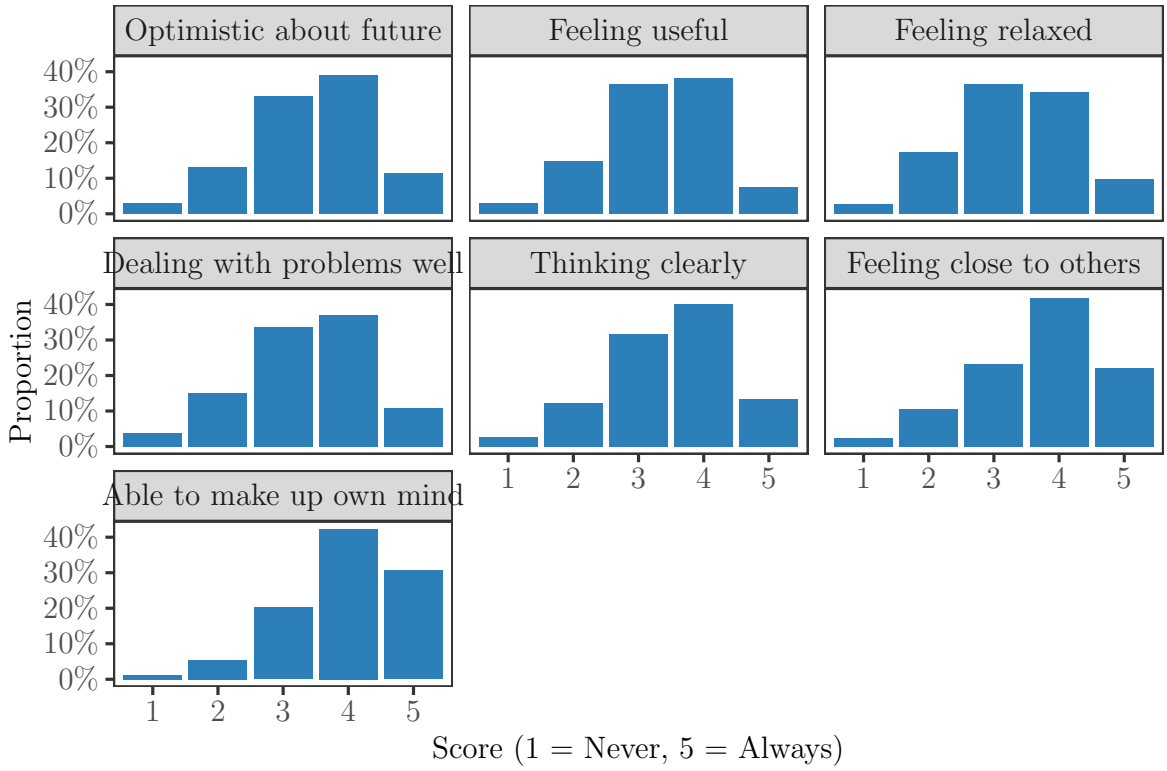
Notes: The table includes observations with non-missing attainment records in both the Millennium Cohort Study (MCS) and the National Pupil Database (NPD) from Stopforth et al. (2025). NPD attainment data are missing for 8.6% of the English language sample and 7.0% of the mathematics sample. Percentages may not sum to 100 due to rounding.

[a] Includes cases in which the grade value matches but was converted to the incorrect grade type.

[b] Includes exact matches and cases converted from a numeric to alphabetical grade which don't align perfectly and are out by one grade at most.

Appendix B. Distribution of individual wellbeing items

Figure B1: Wellbeing individual item distributions



Appendix C. Falsification test - effects of failing English and maths for pupils in Wales and Northern Ireland

Table C1: Resit effects on wellbeing for Wales and Northern Ireland

	Base	Cognitive	Controls	Prior wellbeing
Resit	0.038	−0.005	0.023	0.028
	(0.086)	(0.090)	(0.089)	(0.092)
Observations	1517	1435	1413	1401
R^2	0.019	0.019	0.053	0.139
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls are the same as the main paper but now also include country indicators. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C2: Resit effects on wellbeing for Wales and Northern Ireland by group

	Base	Cognitive	Controls	Prior wellbeing
English resit	0.137	0.074	0.026	−0.004
	(0.138)	(0.139)	(0.144)	(0.136)
Maths resits	−0.014	−0.086	−0.032	−0.021
	(0.170)	(0.192)	(0.162)	(0.167)
Both resits	0.039	0.021	0.076	0.101
	(0.125)	(0.126)	(0.133)	(0.129)
Observations	1517	1435	1413	1401
R^2	0.019	0.020	0.054	0.139
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls are the same as the main paper but now also include country indicators. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C3: Resit effects on wellbeing for Wales and Northern Ireland by group

	Base	Cognitive	Controls	Prior wellbeing
Any resit	0.037	−0.024	−0.009	−0.061
	(0.112)	(0.118)	(0.110)	(0.113)
Any resit × both	0.003	0.049	0.087	0.106
	(0.162)	(0.165)	(0.168)	(0.158)
Observations	1517	1435	1413	1401
R^2	0.019	0.019	0.054	0.131
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls are the same as the main paper but now also include country indicators. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C4: Resit effects on individual wellbeing items for Wales and Northern Ireland

	Optimism	Useful	Thinking clearly	Relaxed	Make mind up	Close to others	Problems
Resit maths	−0.023	0.162	−0.143	−0.043	−0.085	0.015	−0.050
	(0.109)	(0.111)	(0.190)	(0.186)	(0.154)	(0.092)	(0.193)
Resit English	−0.045	−0.212	0.218	0.107	−0.004	0.058	0.127
	(0.105)	(0.142)	(0.140)	(0.134)	(0.118)	(0.118)	(0.135)
Observations	1401	1401	1401	1401	1401	1401	1401
R^2	0.056	0.122	0.114	0.139	0.052	0.100	0.104
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Outcome mean	3.48	3.38	3.55	3.32	4.03	3.82	3.41

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variables are the individual items from the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C5: Resit effects with gender interactions for Wales and Northern Ireland

	Base	Cognitive	Controls	Prior wellbeing
Resit maths × male	0.007	0.002	0.036	−0.006
	(0.140)	(0.148)	(0.152)	(0.148)
Resit English × male	0.033	0.022	0.052	0.046
	(0.122)	(0.123)	(0.130)	(0.124)
Resit maths × female	−0.054	−0.053	−0.032	0.026
	(0.205)	(0.208)	(0.173)	(0.177)
Resit English × female	0.048	0.027	0.073	0.079
	(0.213)	(0.214)	(0.188)	(0.173)
Observations	1435	1435	1413	1401
R^2	0.036	0.037	0.054	0.139
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, sex and country. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C6: Resit effects with SEND interactions for Wales and Northern Ireland

	Base	Cognitive	Controls	Prior wellbeing
Resit maths × no SEND	−0.113	−0.140	−0.059	−0.021
	(0.147)	(0.161)	(0.144)	(0.151)
Resit English × no SEND	0.072	0.080	0.050	0.045
	(0.134)	(0.135)	(0.133)	(0.125)
Resit maths × SEND support	−0.068	−0.026	−0.052	−0.156
	(0.227)	(0.216)	(0.191)	(0.176)
Resit English × SEND support	−0.195	−0.248	−0.237	−0.208
	(0.302)	(0.278)	(0.267)	(0.235)
Resit maths × EHCP	0.165	0.103	0.310	0.007
	(0.471)	(0.480)	(0.453)	(0.242)
Resit English × EHCP	0.350	0.381	0.261	0.267
	(0.532)	(0.544)	(0.500)	(0.419)
Observations	1517	1435	1413	1401
R^2	0.004	0.012	0.046	0.132
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, sex and country. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C7: Resit effects by distance to pass threshold for Wales and Northern Ireland

	Base	Cognitive	Controls	Prior wellbeing
Maths: close	0.086	0.078	0.116	0.097
	(0.179)	(0.195)	(0.167)	(0.171)
Maths: not close	−0.189	−0.287**	−0.168	−0.087
	(0.144)	(0.130)	(0.134)	(0.143)
English: close	0.155	0.165	0.131	0.102
	(0.125)	(0.123)	(0.121)	(0.117)
English: not close	0.077	0.036	−0.005	0.007
	(0.209)	(0.203)	(0.207)	(0.193)
Observations	1517	1435	1413	1401
R^2	0.023	0.026	0.057	0.141
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, sex and country. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C8: Resit effects by month for Wales and Northern Ireland

	Base	Cognitive	Controls	Prior wellbeing
Any resit × Jan/Feb	0.402*	0.250	0.258	0.257
	(0.206)	(0.157)	(0.176)	(0.176)
Any resit × March/Apr	−0.316*	−0.312*	−0.246	−0.181
	(0.167)	(0.174)	(0.166)	(0.156)
Any resit × May/June	−0.071	0.102	0.130	0.096
	(0.129)	(0.115)	(0.118)	(0.114)
Any resit × July on	−0.114	−0.294**	−0.215	−0.203
	(0.128)	(0.130)	(0.131)	(0.145)
Observations	1517	1435	1413	1401
R^2	0.017	0.023	0.051	0.136
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, sex and country. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Appendix D. Timing by subject results

Table D1: Resit effects on wellbeing by month by subject

	Base	Cognitive	Controls	Prior wellbeing
Resit maths × Jan/Feb	−0.713**	−0.245*	−0.262**	−0.221*
	(0.285)	(0.126)	(0.131)	(0.124)
Resit maths × March/Apr	−0.084	−0.036	0.018	−0.010
	(0.089)	(0.096)	(0.096)	(0.087)
Resit maths × May/June	0.082	−0.075	−0.009	−0.033
	(0.309)	(0.150)	(0.141)	(0.132)
Resit maths × July on	0.383	−0.059	0.006	0.049
	(0.341)	(0.195)	(0.225)	(0.167)
Resit English × Jan/Feb	0.347	−0.111	−0.121	−0.096
	(0.345)	(0.114)	(0.113)	(0.100)
Resit English × March/Apr	−0.059	−0.071	−0.139	−0.128
	(0.091)	(0.096)	(0.096)	(0.088)
Resit English × May/June	−0.292	−0.135	−0.190	−0.162
	(0.230)	(0.158)	(0.158)	(0.164)
Resit English × July on	−0.473	0.058	−0.011	0.114
	(0.307)	(0.231)	(0.263)	(0.185)
Observations	4776	4464	4391	4358
R^2	0.020	0.008	0.052	0.145
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, sex, SEND status and month of interview. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Appendix E. Robustness check - interacting resitting English and maths

Table E1: Main results robustness - interaction effect

	Base	Cognitive	Controls	Prior wellbeing
Resit maths	0.114	−0.163**	−0.137*	−0.095
	(0.209)	(0.079)	(0.077)	(0.076)
Resit English	0.059	−0.155*	−0.218**	−0.134*
	(0.196)	(0.086)	(0.092)	(0.076)
Resit maths × Resit English	−0.389	0.201	0.234*	0.115
	(0.294)	(0.136)	(0.134)	(0.124)
Observations	4776	4464	4391	4358
R^2	0.005	0.007	0.052	0.143
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. The table reports interactions between GCSE resit indicators and sex. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Appendix F. Robustness check - flag for failing any other exams

The following models all include flags for whether the student failed any other GCSEs (either no fails, one other fail or two or more other fails).

Table F1: Main results robustness - failed any exams

	Base	Cognitive	Controls	Prior wellbeing
Resit	0.084	−0.094	−0.085	−0.062
	(0.125)	(0.060)	(0.059)	(0.054)
Observations	4764	4453	4380	4347
R^2	0.006	0.007	0.053	0.143
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. The table reports interactions between GCSE resit indicators and sex. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table F2: Main results robustness by subject - failed any exam

	Base	Cognitive	Controls	Prior wellbeing
Resit maths	0.031	−0.054	0.000	−0.022
	(0.148)	(0.067)	(0.070)	(0.062)
Resit English	−0.040	−0.032	−0.067	−0.043
	(0.175)	(0.073)	(0.075)	(0.067)
Observations	4764	4453	4380	4347
R^2	0.005	0.007	0.052	0.143
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. The table reports interactions between GCSE resit indicators and sex. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table F3: Gender results robustness - failed any exam

	Base	Cognitive	Controls	Prior wellbeing
Resit maths × male	0.133	0.115	0.146	0.088
	(0.104)	(0.103)	(0.108)	(0.096)
Resit English × male	−0.208**	−0.235**	−0.190*	−0.126
	(0.100)	(0.097)	(0.102)	(0.084)
Resit maths × female	−0.114	−0.130	−0.125	−0.115
	(0.083)	(0.085)	(0.087)	(0.083)
Resit English × female	0.032	0.013	0.077	0.052
	(0.105)	(0.105)	(0.104)	(0.101)
Observations	4453	4453	4380	4347
R^2	0.041	0.044	0.054	0.144
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. The table reports interactions between GCSE resit indicators and sex. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table F4: SEND results robustness - failed any exam

	Base	Cognitive	Controls	Prior wellbeing
Resit maths × no SEND	0.012	−0.091	−0.037	−0.051
	(0.163)	(0.070)	(0.075)	(0.064)
Resit English × no SEND	−0.020	−0.005	−0.057	−0.038
	(0.190)	(0.080)	(0.080)	(0.071)
Resit maths × SEND support	0.372**	0.355*	0.354*	0.169
	(0.187)	(0.189)	(0.190)	(0.208)
Resit English × SEND support	−0.031	−0.059	−0.045	−0.017
	(0.175)	(0.182)	(0.184)	(0.184)
Resit maths × EHCP	0.181	0.180	0.190	0.270
	(0.192)	(0.196)	(0.211)	(0.254)
Resit English × EHCP	−0.293	−0.329	−0.308	−0.227
	(0.200)	(0.201)	(0.221)	(0.230)
Observations	4764	4453	4380	4347
R^2	0.009	0.010	0.054	0.144
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. The table reports interactions between GCSE resit indicators and sex. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table F5: Month of exam results robustness to failing any exam

	Base	Cognitive	Controls	Prior wellbeing
Any resit × Jan/Feb	0.105	−0.135	−0.126	−0.103
	(0.254)	(0.105)	(0.102)	(0.091)
Any resit × March/Apr	0.001	−0.064	−0.054	−0.057
	(0.092)	(0.079)	(0.081)	(0.072)
Any resit × May/June	0.057	−0.122	−0.098	−0.100
	(0.206)	(0.124)	(0.117)	(0.107)
Any resit × July on	0.201	−0.068	−0.082	0.055
	(0.260)	(0.134)	(0.147)	(0.120)
Observations	4764	4453	4380	4347
R^2	0.011	0.008	0.053	0.144
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. The table reports interactions between GCSE resit indicators and sex. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table F6: Resit effects on wellbeing for close vs not close fails robustness to failing any exam

	Base	Cognitive	Controls	Prior wellbeing
Maths: close	0.061	−0.082	−0.024	−0.039
	(0.159)	(0.076)	(0.077)	(0.066)
Maths: not close	0.033	0.032	0.077	0.042
	(0.205)	(0.108)	(0.107)	(0.096)
English: close	0.019	−0.015	−0.046	−0.019
	(0.182)	(0.073)	(0.074)	(0.068)
English: not close	−0.220	−0.121	−0.164	−0.142
	(0.191)	(0.128)	(0.138)	(0.122)
Observations	4764	4453	4380	4347
R^2	0.007	0.007	0.053	0.144
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, sex and country. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table F7: Resit effects on wellbeing for close vs not close fails by gender robustness to failing any exam

	Base	Cognitive	Controls	Prior wellbeing
Mathematics				
Maths: close × male	0.146	0.141	0.168	0.139
	(0.120)	(0.117)	(0.123)	(0.104)
Maths: close × female	−0.157	−0.169*	−0.162	−0.165*
	(0.099)	(0.100)	(0.101)	(0.090)
Maths: not close × male	0.183	0.153	0.182	0.074
	(0.143)	(0.142)	(0.145)	(0.134)
Maths: not close × female	−0.015	−0.037	−0.042	0.004
	(0.147)	(0.150)	(0.145)	(0.129)
English				
English: close × male	−0.155	−0.179*	−0.138	−0.073
	(0.099)	(0.096)	(0.098)	(0.089)
English: close × female	0.010	−0.006	0.051	0.026
	(0.118)	(0.118)	(0.115)	(0.099)
English: not close × male	−0.364**	−0.396**	−0.343*	−0.269**
	(0.173)	(0.167)	(0.177)	(0.134)
English: not close × female	0.000	−0.021	0.072	0.014
	(0.226)	(0.225)	(0.226)	(0.246)
Observations	4453	4453	4380	4347
R^2	0.043	0.045	0.055	0.145
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, sex and country. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table F8: Resit effects on individual wellbeing items - robustness to including failed any exam group

	Optimism	Useful	Thinking clearly	Relaxed	Make mind up	Close to others	Problems
Resit maths	−0.044	−0.063	−0.137*	0.006	0.018	0.081	−0.031
	(0.086)	(0.083)	(0.078)	(0.062)	(0.074)	(0.083)	(0.113)
Resit English	−0.037	−0.054	−0.008	0.053	−0.115	−0.089	0.104
	(0.091)	(0.069)	(0.066)	(0.062)	(0.080)	(0.074)	(0.106)
Observations	4347	4347	4347	4347	4347	4347	4347
R^2	0.060	0.107	0.122	0.138	0.056	0.071	0.097
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Outcome mean	3.43	3.33	3.50	3.32	3.97	3.71	3.37

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. TThe dependent variables are the individual items from the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table F9: Resit effects on individual wellbeing items - gender interactions. Robustness to including failed any exam group

	Optimism	Useful	Thinking clearly	Relaxed	Make mind up	Close to others	Problems
Resit maths - Male	−0.001	0.110	0.063	0.139	0.030	0.235	−0.129
	(0.104)	(0.121)	(0.092)	(0.085)	(0.109)	(0.149)	(0.214)
Resit maths - Female	−0.076	−0.206**	−0.293**	−0.095	0.011	−0.042	0.045
	(0.106)	(0.098)	(0.113)	(0.092)	(0.094)	(0.072)	(0.101)
Resit English - Male	−0.086	−0.153	−0.092	−0.002	−0.127	−0.178	0.132
	(0.099)	(0.093)	(0.069)	(0.070)	(0.090)	(0.111)	(0.169)
Resit English - Female	0.009	0.054	0.060	0.090	−0.107	0.000	0.090
	(0.154)	(0.113)	(0.114)	(0.099)	(0.134)	(0.075)	(0.116)
Observations	4347	4347	4347	4347	4347	4347	4347
R^2	0.059	0.109	0.125	0.140	0.056	0.073	0.098
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Outcome mean	3.43	3.33	3.50	3.32	3.97	3.71	3.37

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variables are the individual items from the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Appendix G. Robustness check - resit effects on individual items by month group

Table G1: Resit effects on individual wellbeing items by month

	Optimism	Useful	Thinking clearly	Relaxed	Make mind up	Close to others	Problems
Resit maths × Jan/Feb	−0.555***	−0.233*	−0.184	0.057	−0.083	−0.088	−0.009
	(0.147)	(0.119)	(0.135)	(0.152)	(0.185)	(0.138)	(0.133)
Resit maths × March/Apr	0.056	−0.154*	−0.166	0.010	−0.032	0.009	0.087
	(0.125)	(0.092)	(0.102)	(0.089)	(0.109)	(0.084)	(0.103)
Resit maths × May/June	−0.004	−0.109	−0.168	−0.069	0.044	0.105	−0.072
	(0.151)	(0.176)	(0.199)	(0.118)	(0.095)	(0.116)	(0.168)
Resit maths × July on	0.060	0.245	−0.134	0.044	0.123	0.401	−0.486
	(0.161)	(0.219)	(0.150)	(0.173)	(0.261)	(0.253)	(0.368)
Resit English × Jan/Feb	0.041	−0.133	−0.210**	0.038	−0.293**	0.037	−0.086
	(0.133)	(0.107)	(0.104)	(0.123)	(0.131)	(0.113)	(0.107)
Resit English × March/Apr	−0.082	−0.057	−0.098	−0.003	−0.118	−0.132	−0.001
	(0.131)	(0.087)	(0.086)	(0.094)	(0.119)	(0.093)	(0.106)
Resit English × May/June	−0.204	−0.208	−0.118	0.061	−0.207	−0.055	−0.072
	(0.166)	(0.163)	(0.178)	(0.131)	(0.139)	(0.101)	(0.175)
Resit English × July on	−0.151	−0.051	0.420***	0.124	0.130	−0.300	0.528*
	(0.134)	(0.219)	(0.156)	(0.151)	(0.247)	(0.252)	(0.276)
Observations	4358	4358	4358	4358	4358	4358	4358
R^2	0.063	0.110	0.126	0.139	0.059	0.074	0.100
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Outcome mean	3.43	3.33	3.50	3.32	3.97	3.71	3.37

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variables are the individual items from the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Appendix H. Restricted comparison group analysis

Table H1: Effects of any resit on wellbeing: restricted comparison sample

	Base	Cognitive	Controls	Prior wellbeing
Resit	−0.144	−0.188***	−0.201***	−0.176***
	(0.115)	(0.065)	(0.066)	(0.059)
Observations	2114	1951	1916	1888
R^2	0.004	0.010	0.061	0.148
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table H2: Effects of resit group on wellbeing: restricted comparison sample

	Base	Cognitive	Controls	Prior wellbeing
English resit	−0.032	−0.182*	−0.250**	−0.188**
	(0.205)	(0.095)	(0.102)	(0.085)
Maths resits	0.023	−0.210**	−0.166*	−0.147*
	(0.222)	(0.089)	(0.085)	(0.081)
Both resits	−0.307***	−0.178**	−0.190**	−0.192**
	(0.108)	(0.089)	(0.092)	(0.086)
Observations	2114	1951	1916	1888
R^2	0.014	0.010	0.061	0.148
Cognitive test scores		✓	✓	✓
Demographics			✓	✓
Prior wellbeing				✓

* p <0.1, ** p <0.05, *** p <0.01

Notes: Survey-weighted estimates are reported. Survey-design-based standard errors in parentheses account for the stratified and clustered MCS sampling design. The dependent variable is the standardised Warwick–Edinburgh wellbeing scale. Cognitive test scores include a Word Activity cognitive test (age 14) and a Number Analogies test (age 17). Demographic controls include household income, household size, ethnicity, SEND status, month of interview and sex. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

ucl.ac.uk/ioe/cepeo