Lessons learnt from EEF Post-16 evaluations: Recommendations for evaluators

September 2024

I. About this paper

The socioeconomic attainment gap grows as learners progress through the education system, meaning that it is at its widest when they reach the post-16 stage. In 2023, by the end of secondary school, socioeconomically disadvantaged pupils were 18.8 months behind their peers, whilst persistently disadvantaged pupils (those eligible for free school meals (FSM) for at least 80 per cent of their time at school) were almost two years (22.7 months) behind (EPI, 2023). The post-16 stage is therefore a 'last chance' to try to minimise the socioeconomic attainment gap and improve life chances before most young people leave the education system.

In 2023, the Department for Education (DfE) announced an additional £40 million of funding for the Education Endowment Foundation (EEF) to further build the evidence base and support effective practice in the post-16 sector. This additional funding will allow the EEF to continue generating evidence on supporting GCSE resits, which has been the focus of its work to date, while expanding into other academic, technical and vocational study programmes[1] and working with the sector to answer key questions through robust research.

This paper draws on the EEF's early experience of funding and managing post-16 interventions aimed at supporting GCSE resits and their evaluations, relevant research commissioned by the EEF, as well as feedback received from evaluators, developers and other researchers in the post-16 sector. While we have faced various challenges, the scarcity of rigorous research at post-16 and the importance of this stage of education as a 'last chance' for supporting socioeconomically disadvantaged learners means there are significant benefits for educators and policy makers if we can tackle them. To that end, this paper is a first attempt at providing a set of focussed, practical recommendations for teams planning to conduct post-16 evaluations for the EEF. It is intended as an early summary of lessons learnt from past evaluations and does not aim to provide an exhaustive list of risks or mitigation strategies.

II. Background

This section outlines some background information on the post-16 sector, including the types of setting and qualification, the learner enrolment process and transition to post-16 and other relevant context, as well as the EEF's experience commissioning post-16 evaluations to date.

Types of post-16 setting

Post-16 education in England comprises adolescents aged 16-19 who are continuing their education or receiving vocational training (either full- or part-time)[2]. In this sector, young people are referred to as 'learners' (not 'pupils') and the providers of their education are referred to as 'colleges' (not 'schools', though some secondary schools do provide post-16 education), though the terms 'students' and 'settings', respectively, can also be used. In this document, 'setting' will be used to refer to post-16

[1] Note that work-based qualifications (e.g., apprenticeships) currently sit beyond the scope of the EEF's post-16 work.

[2] Adolescents can leave *school* when they turn 16 (which is when most pupils will write their GCSEs at the end of KS4). However, it is compulsory for children to stay in *education* until they are 18. Between 16 and 18, children can stay in full-time education (e.g., at a college), start an apprenticeship or traineeship or spend 20 hours or more a week working or volunteering, while in part-time education or training.

education providers, which can in turn comprise multiple campuses. There are **three main types of setting offering post-16 education** and provision in these different contexts varies:

- **School sixth form**. These are secondary schools that integrate Year 12 and Year 13 into their provision, and usually offer a variety of A levels and some other qualifications to learners aged 16-19 years old. These types of settings are more likely to have an academic slant and employ qualified teachers. While they provide education to 41% of all post-16 learners, a relatively small proportion of these learners are from socio-economically disadvantaged backgrounds or low attaining. Note that 'sixth form' is the term used to collectively describe Year 12 and Year 13 groups (i.e., Key Stage 5), which are not part of the national curriculum.
- **Sixth form college**. These are colleges that offer Year 12 and Year 13 provision to learners aged 16-19 years old, but that are not attached to a secondary school. They have a more informal learning environment than school sixth forms, where learners take responsibility for themselves and their learning, and tend to offer a greater variety of qualifications. Like school sixth forms, they are more likely to have an academic slant and employ qualified teachers, and less likely to serve socio-economically disadvantaged or low attaining learners. They represent the smallest proportion of overall post-16 provision, both in terms of the number of settings and proportion of learners served.
- **Further education (FE) college**. These are colleges open to *any age group* and tend to offer the greatest variety of academic and vocational qualifications, as well as different study options (e.g., part-time, distance learning). They may also offer apprenticeships or courses with work placements. FE colleges are usually quite large and dispersed, often operating across several campuses. While there are only 156 general FE colleges in England, these comprise approximately 800 sites or campuses (estimate provided by the Association of Colleges). Despite only accounting for 6% of the total number of settings, FE colleges serve the largest proportion of learners, particularly those that are socio-economically disadvantaged, as can be seen in Table 1. Relative to other setting types, disadvantage funding makes up a greater proportion of FE colleges' budgets, driven primarily by learners' low prior attainment (block 2) rather than their deprivation level (block 1 – see 'Funding to support disadvantaged learners' section below). Compared to other types of settings, FE colleges have more of a vocational slant and are more likely to employ non-specialist English and maths teachers.

Note that there are other types of setting in the sector that offer mainstream vocational education and training to young people and adults. While the largest of these 'other' types by number of settings is Special Academy (7% of settings), the largest by number of learners (3%) is Independent Training Providers (ITPs). ITPs are distinct from other types of FE setting in that they are not run or directly controlled by the state. Furthermore, the majority of qualifications offered by ITPs are work-based (e.g., apprenticeships), though some do offer academic and technical qualifications.

Table 1: Distribution of settings and learners by type of provider

Type of provider	Settings		Total learners		% disadvantage funding		
	N	%	N	%	Block 1	Block 2	Total
School Sixth Form	1,937	69%	492,113	41%	2.8%	1.2%	4.0%
Sixth Form College	44	2%	104,221	9%	3.8%	1.9%	5.7%
FE College	156	6%	504,980	43%	4.4%	6.3%	10.7%
Other	685	24%	85,846	7%	2.9%	5.4%	8.4%
Total	**2,822**	**-**	**1,187,159**	**-**	**4.0%**	**3.6%**	**7.7%**

Notes: (1) The information in this table is based on 16-19 funding allocation data from the 2023-24 academic year. (2) 'The 'School Sixth Form' category also includes Academy, Free School and Free School 16-19. (3) The 'Other' category includes: Alternative Provision (Academy and Free School), Special (Academy, Free School, Non-Maintained School and Post-16 Institution), Agricultural & Horticultural College, Art & Design College, City technology college, Higher Education Provider, Independent Learning Provider, Local Authority, Non-Maintained Special School, Specialist Designated College, Studio School, University Technical College and Other. (4) '% disadvantage funding' refers to the proportion of settings' total funding that is categorised as 'disadvantage funding'. Disadvantage funding is based on learners' economic deprivation (block 1) and

low prior attainment in English and maths (block 2). If a setting's total disadvantage funding (block 1 plus block 2) is less than £6,000, it is topped up to this value. This explains why the 'Block 1' and 'Block 2' columns above do not always add up to the 'Total' column, with the 'Top-up' column being '0%' for Sixth Form and FE colleges.

Post-16 qualifications and GCSE resits

Of the nine qualification levels in England set out by the Regulated Qualifications Framework (RQF), **most 16-18-year-olds are working towards their entry level qualification (Level 1, Level 2 or Level 3)**. While these qualification *levels* are relatively stable, the *types* of qualifications have seen a lot of change over time[3]. This results in thousands of qualifications being on offer, all with different learning aims and lengths of study (e.g., qualifications labelled as awards, diplomas and certificates that depend on the number of study hours completed) and offered by different awarding bodies (e.g., BTECs by Pearson, Cambridge Technicals by OCR, etc). Nonetheless, the **most common qualifications can be organised into three main types: academic, technical and work-based qualifications**[4].

Table 2: Entry level qualifications, by level and type

	Academic qualifications Studying more than one subject in depth	**Technical qualifications** Taking vocational courses that combing classroom learning and practical activities	**Work-based qualifications** Developing knowledge and skills for a particular job through practical learning in the workplace and with a training provider
Level 3	A level, AS level, international baccalaureate diploma	T level, applied general (e.g., BTEC Nationals, Level 3 Cambridge Technicals), level 3 certificate/ diploma/ award	Advanced apprenticeship, level 3 NVQ
Level 2	GCSE grades 4-9 (A*-C)	Level 2 certificate/ diploma/ award, level 2 functional skills	Intermediate apprenticeship; level 2 NVQ
Level 1	GCSE grades 1-3 (D-G)	Level 1 certificate/ diploma/ award, level 1 functional skills	Traineeship, supported internship, pre-apprenticeship, level 1 NVQ

Note: Despite being grouped as 'technical' qualifications, T-levels and applied general qualifications (e.g., BTEC Nationals, Level 3 Cambridge Technicals) can be used to progress to higher education. As is the case with A-levels, the qualification and grade achieved is converted into a numerical value (UCAS points), which Higher Education providers use to assess whether learners meet their entry requirements.

At Level 2, the qualification that receives the most attention is **GCSE grades**, particularly achieving a grade 4 or above (a grade C or above in the previous grading system) in English and maths. Studies show that achieving a grade 4 (C) in these exams, which is referred to as a 'standard pass', is correlated with better academic, career and other future outcomes. This is likely why, as of September 2014, policy mandates that 16-year-olds who don't get at least a grade 4 (C) in their GCSE English and maths are required to keep on studying the subjects until they are 18 or secure a GCSE grade 4 (C) or above in

[3] Most recently, the previous government removed funding for qualifications with low or no enrolment and initiated a review of Level 3 technical qualifications to remove funding for those which overlap with T-levels. In the longer term, it had planned to transition to the Advanced British Standard (ABS), a baccalaureate-style qualification that would bring academic and technical pathways together into a single framework, and allow learners to study a technical route based on the T-Level model, academic subjects based on the A-level model, or a combination of both. The new government has committed to pause and reviewthe qualification reforms and has cancelled the transition to the ABS. Instead, it plans to establish Skills England to bring together central and local government, businesses, training providers and unions to meet the skills needs of the next decade across all regions, providing strategic oversight of the post-16 skills system aligned to the government's Industrial Strategy.

[4] While the EEF's work to date has focused on Level 2 academic qualifications (e.g., achieving a GCSE grade 4 or C), it's expanded remit covers other academic and technical qualifications. Work-based qualifications currently sit beyond the scope of the EEF's post-16 work.

these subjects. That is, learners starting a new study programme with a GCSE grade 3 (equivalent to a D in the previous grading system) in English and/or maths must enrol in GCSE resit courses or Functional Skills Level 2 qualification (see below). This requirement, commonly referred to as the **'GCSE resit policy'** is a **'condition of funding'**: post-16 providers with learners who do not meet this condition (i.e., have a GCSE grade 3 (D) and are not enrolled on the relevant English[5] or maths course), will receive a funding penalty in a future year[6].

Note that, as an alternative to retaking the GCSE, learners who previously achieved a grade 2 (roughly equivalent to a grade E in the previous grading system) or below can study towards a Functional Skills Level 2 qualification[7]. Functional Skills Qualifications (FSQs) were introduced in 2007 as a way for learners to progress into employment or further technical education by developing skills relevant to everyday life and the workplace[8]. Unlike the GCSEs, FSQs are only offered in three subjects (English, maths and digital skills), require fewer hours of study and are graded with a simple pass/fail.

According to data from the Joint Council for Qualifications (Table 3), the rates at which post-16 learners in England achieve a grade 4 on their GCSE resits (hereafter referred to as 'pass rates') are low, especially for maths and for those who take resits in the summer. For reference, 16-year-olds sitting their GCSEs for the first time in 2024 had pass rates of over 70%. The pass rate for post-16 learners who resat GCSE English language exams in summer 2024 was 19.3%, while the figure was 15.0% for GCSE maths. Post-16 learners who sat their exams in November 2023 had substantially higher pass rates: 40.3% for GCSE English language and 22.2% for GCSE Maths. These higher pass rates in November potentially reflect the fact that learners enrolled in mid-year resits are typically closer to passing and will have scored higher than other resitters on their first attempts.

Table 3: GCSE results for English language and maths, 2022/23 to 2023/24

Subject	Year	Learners sitting GCSEs in summer (16-year-olds)		Learners sitting GCSEs in summer (17-19-year-olds)		Learners sitting GCSEs in November (17-19-year-olds)	
		N	Pass %	N	Pass %	N	Pass %
English language	2023/24	630837	71.2%	133411	19.3%	53688	40.3%
	2022/23	607163	71.6%	101980	23.0%	39119	37.6%
Maths	2023/24	628206	72.0%	165796	15.0%	57773	22.2%
	2022/23	605646	72.3%	134653	13.3%	46995	24.2%

Source: Joint Council for Qualifications

The condition of funding has resulted in a lot more learners taking English and maths qualifications at Level 2 and below, especially considering that around 37% of learners require two or more retake attempts to achieve the required GCSE grade (Rodeiro 2018). Furthermore, in February 2024, the DfE

[5] It is only necessary for learners to retake English Language if they did not previously achieve a grade 4 (C) in both English Language and English Literature. That is if they previously achieved a grade 4 (C) in English Literature but not English Language, they do not have to retake English Language.

[6] Note that there are separate condition of funding adjustments for English and maths, meaning that learners who do not have either GCSE will be counted twice.

[7] Learners who previously achieved a grade 3 (C) only have the option of studying towards a grade 4-9 (D-A*) in their GCSE.

[8] Despite being positioned as a more practical alternative to the GCSEs, a 2019 reform to FSQs has made them more similar to GCSEs and, according to feedback from the sector, too difficult for the learners they were intended for (AELP 2024). As a result of this and limited understanding of FSQs by employers, entries to FSQs are now declining, with a 45% reduction in the 2021/22 academic year (Ofqual 2022).

amended the condition of funding for GCSE resits to mandate a minimum of three in-person taught hours a week for English, and four for maths.

Funding to support disadvantaged learners

While settings are penalised for any learners that do not meet the condition of funding, they currently do not receive any targeted funds for supporting them. However, a recent change to the English and maths element of 16-19 funding is intended to provide settings with additional resources to support learners with low prior attainment in these subjects.

In terms of funding to support disadvantaged learners in particular, it should be noted that the Pupil Premium received by state-funded schools in England does not extend to post-16 providers. Instead, settings receive **disadvantage funding** to attract, retain and support disadvantaged learners and those with learning difficulties and disabilities. Disadvantage funding is comprised of two blocks: one to account for learners' economic deprivation (Block 1) and the other to account for low prior attainment in English and maths (Block 2). Block 1 funding is calculated using the Index of Multiple Deprivation (IMD)[9], and includes additional funding for learners who are in/have recently left care and learners living in the 27% most deprived areas of the country. Block 2 funding is calculated using attainment in GCSE English and/or maths at the end of year 11[10] and varies depending on the learners' programme of study. As with other core factors of the 16-19 funding formula, disadvantage funding is not ringfenced, meaning providers are free to choose how to spend this additional funding to support disadvantaged learners.

Additionally, there are two schemes in the student financial support element of 16-19 funding that address disadvantage, both of which are ringfenced unlike disadvantage funding. **FE free meals** are the equivalent of FSM in schools. However, the electronic Eligibility Checking System (ECS) that enables local authorities to check FSM eligibility on behalf of schools cannot currently be used to check post-16 learners' eligibility, and learners (or a parent or guardian on their behalf) must apply directly to the setting at which they are enrolled (either using paper or virtually, depending on the setting). **Discretionary bursaries** support learners with their day-to-day costs of participating in education (e.g., transportation, books and equipment, etc). The eligibility criteria for these bursaries, and the amount awarded to eligible learners, are at the discretion of the setting.

There is also 'high needs funding' for settings with learners who are assessed by their local authority as having complex special educational needs and disabilities (SEND). Settings receive extra funding directly from the local authority to cover the costs of additional support in excess of £6,000 for these learners.

Learner enrolment and transition to post-16[11]

Pupils can apply to several colleges and courses in their final year of secondary school, and consequently can have several conditional offers from which to choose once they receive their GCSE results at the end of August. While they can then enrol in their preferred option, they don't have to formally do so for the first six weeks of term. Many learners also arrive as 'walk ins' on enrolment day

---

[9] Eligibility for Block 1 funding is determined by learners' home postcode and the level of deprivation recorded in the IMD for that area. The IMD is published by the Department for Levelling Up, Housing and Communities, and the most recent version (currently IMD 2019) is used to determine Block 1 funding.

[10] For Block 2 funding, low prior attainment is defined as not having achieved a grade 4 or above in GCSE English and/or maths by the end of year 11. Learners who have achieved an English literature GCSE, but not English language, are eligible for block 2 funding. This is different from the condition of funding criteria, where both English language and literature count as positive prior attainment. Note that each instance of low attainment in English and maths is counted separately, meaning that learners who do not have either GCSE will be counted twice.

[11] Note that the information in this section is anecdotal, having been compiled from conversations with various stakeholders in the post-16 sector.

once they've received their exam results. These numbers tend to be lower for sixth forms, but FE colleges can see as much as 50% of the cohort arriving on the day without a prior application.

During the enrolment period, learners may leave one college and/or course and start another. Furthermore, many colleges take a 'swap don't drop' approach to courses to encourage learners to stay enrolled, which leads to a lot of movement between courses within the first month or so of enrolment. This means that colleges won't have a final list of registered learners until mid- to late October.

In addition, colleges often do not have ready access to information on those learners' prior attainment and individual needs. Often, little information is accessible from schools (though school sixth forms can usually access information from the school they are attached to) and the data that can be accessed through administrative sources is limited. As a result, colleges often spend the first six weeks of term using diagnostic assessments to identify learners' strengths and weaknesses so they can tailor their support accordingly. The quality of these assessments is unclear, as is their alignment with the assessments for post-16 qualifications and prior key stages.

There is currently no existing system for passing information automatically from schools to colleges. Some colleges, particularly sixth forms within the same school group, get around this by developing strong relationships with feeder schools. Generally, once learners have left school, capacity is focused on the incoming cohort, which does not leave much time for data being sent to colleges. Colleges therefore often re-collect all the information schools already hold, including previous FSM eligibility and SEND needs. Colleges may hire interns on enrolment day to manage the data entry. Colleges may struggle to find the right Unique Learner Number (ULN) for learners, for example because they are going by a slightly different name. For some colleges this happens with up to 40% of learners.

Other helpful sector context

- The data that are available in administrative sources is more limited at post-16. While the DfE's National Pupil Database (NPD) consolidates information on basic learner demographics and learning aims, post-16 settings differ in how they report data about their learners according to type. School sixth forms continue to report learner data through the school census (although only basic demographic information and learning aims, not FSM status, attendance, etc), while FE and sixth form colleges report learner data through the Individualised Learner Record (ILR).
- Whereas the Unique Pupil Number (UPN) is the DfE's unique identifier for school-aged children in the NPD, the ULN is the unique identifier for post-16 learners, used in both the NPD and ILR. ULNs are generated by the Learning Records Service from when a child turns 14 years of age.
- While term dates for school sixth forms and sixth form colleges mirror those of secondary schools, FE colleges tend to end the academic year in June. This means that staff often take holidays in July rather than August.
- Standard practice for involving young people in research is to have them provide their own consent for participating from the age of 16. While the age of majority in England is 18 years, the Mental Capacity Act defines 'adult' as a person aged 16 years or older and many research ethics committees accept the age of 16 for sole consent on the part of the young person (particularly for low-risk research). This also applies to data protection legislation and, as such, post-16 learners who participate in EEF projects can exercise *their own* rights over their personal data (rather than it being the responsibility of their parents/carers). It is important to note that children and young people under the age of 18 still have a legal right to safeguarding, and researchers must ensure that 16- and 17-year-olds are given the same protection and entitlements as any other child.
- In addition to the DfE, key players in the post-16 sector include:
    - **Education and Skills Funding Agency (ESFA)**, the executive agency of the DfE responsible for funding education and skills providers in England.

- **Office of Qualifications and Examinations Regulation (Ofqual)**, the non-ministerial department of government that regulates qualifications, examinations and assessments in England.
- **Association of Colleges (AoC)**, a not-for-profit membership organisation representing the interests of further education, sixth form, tertiary and specialist colleges in England.
- **Education and Training Foundation (ETF)**, the main workforce development body for the further education and training sector, partnering with others to provide professional development for teachers, trainers and leaders of learners aged 14 and over. It recently delivered the Centres for Excellence in Maths (CfEM) project, which focussed particularly on a teaching for mastery programme adapted from the National Centre for Excellence in the Teaching of Mathematics (NCETM). Though the project ended in 2023, NCETM is currently moving into the post-16 sector with their own professional development programme: Post-16 GCSE and FSQ Mastery Specialist Programme.
- **Awarding bodies** (also referred to as exam boards), the entities responsible for setting and marking examinations and awarding qualifications. The eight largest awarding bodies in the UK are AQA, CCEA (Northern Ireland Council for Curriculum, Examinations and Assessment), City & Guilds, NCFE, OCR, Pearson, SQA (Scottish Qualification Authority) and WJEC (Welsh Joint Education Committee). They are represented by the Joint Council for Qualifications (JCQ).

III. Key challenges and recommendations

As mentioned above, the EEF has funded a few programmes aimed at supporting GCSE resits (see Appendix 1 for a summary) and consolidate some early lessons learnt from their evaluations. To date, the key challenges to running trials in the post-16 sector tend to concentrate around designing and implementing studies with sufficient power to detect a meaningful impact for a given intervention, from the sensitivity of outcome measures and difficulties of accessing/collecting data, to a more fragmented pool of potential participants and their limited interest in engaging with research as well as 'noisier' business as usual. We recommend the following aspects are considered carefully when planning and completing evaluations at post-16, given our aspirations of conducting high-quality evaluations to fill the existing evidence gaps in the sector.

Primary outcome measure and accessing data

To date, the primary outcome of interest in EEF post-16 evaluations has been attainment in English or maths, as measured by GCSE results. Most GCSEs, including English and Maths, now follow a linear system where all assessments are sat in May-June at the end of two years of teaching[12]. Exam boards derive GCSE grades from learners' raw marks based on exam difficulty and the distribution of learners' performance. GCSE grades range from 1 to 9, a 9 being equivalent to a high A* under the previous grading system, and a 1 being equivalent to a G. The lowest grade is a U, meaning ungraded. A grade 4 is considered the standard for a 'pass'.

GCSE English is typically split into English Language and English Literature. GCSE English Language focuses on reading, writing, listening, and speaking skills, and learners are required to retake this subject if they do not achieve a grade 4 on the exam. There are two exam papers for GCSE English Language, plus a Spoken Language Endorsement (which does not count towards grades). AQA has two 105-minute exams (Exploration in Creative Reading & Writing; Writer's Viewpoints and Perspectives) each worth 50% (80 marks each). OCR has two 120-minute exams (one on non-fiction, one on literary texts) each worth 50% (80 marks each). Pearson has one 105-minute exam (Fiction &

[12] Previously, GCSEs followed a modular structure, where assessments were sat at multiple time points, typically over a two-year period of teaching. A Uniform Mark Scale system was used to ensure comparability of grades between years.

Imaginative Writing) worth 64 marks (40%) and a 125-minute exam (Non-Fiction & Transactional Writing) worth 96 marks (60%).

GCSE maths covers six topics: (1) number, (2) algebra, (3) ratio, proportion and rates of change, (4) geometry and measures, (5) probability and (6) statistics. It is assessed through three 90-minute exams (2 calculator, 1 non-calculator). AQA and Pearson papers are 80 marks each, while OCR papers are 100 marks. GCSE maths learners are assigned to either a Higher or Foundation tier paper, depending on the grade they are aiming to achieve. Learners assigned to the higher tier can achieve a grade 4-9, with a 'safety net' of 3 for those who just miss out on a 4. Learners assigned to the Foundation tier can achieve a grade 1-5. While some questions may be the same on both papers, there is a different focus to the content for each.

GCSE grades and pass/fail indicators using the grade 4 (C) threshold are the most policy relevant attainment outcome for post-16 research focused on GCSE resit students. As the same grades are used each year, they are a fairly consistent indicator for measuring attainment over time. However, because of their banded nature, they are not a very sensitive measure for monitoring progress. That is, while a learner moving from the lower end of one GCSE grade distribution to just under the threshold for the next GCSE grade would have improved, this improvement would be masked by not having achieved that next grade.

This means that, to measure the same treatment effect, a greater sample size is needed when using GCSE grades (i.e., there is a loss of power when GCSE grades are used instead of GCSE marks). The indicative power calculations in Table 5 suggest that at least 2.5 times as many participants would be needed to detect the same effect size for maths, while the difference is smaller for English. For a further discussion on this topic please see the guidance on using GCSE as a performance measure (Smith et al, 2021).

Table 5: Power and sample size implications of using GCSE grades vs marks

MDES	Maths			English		
	Power to detect an effect		Ratio of sample size needed to detect an effect (marks:grades)	Power to detect an effect		Ratio of sample size needed to detect an effect (marks:grades)
	Marks	Grades		Marks	Grades	
0.05	0.801	0.655	3.189	0.801	0.739	1.187
0.10	0.798	0.714	2.756	0.799	0.746	1.146
0.15	0.800	0.737	2.607	0.811	0.770	1.082
0.20	0.804	0.750	2.556	0.800	0.785	1.029
0.25	0.799	0.752	2.531	0.802	0.793	0.995

As a result, GCSE *marks*, the raw data underlying the final grade, have recently been used as the primary outcome measure in EEF-funded evaluations (to date, the 5Rs approach to GCSE Maths resits and Maths-for-Life). Unfortunately, GCSE marks are not comparable between awarding bodies (or over time[13]) because they use different schemes for marking scripts. So before using GCSE marks for analysis, these need to be standardised using the mean score and standard deviation of the population

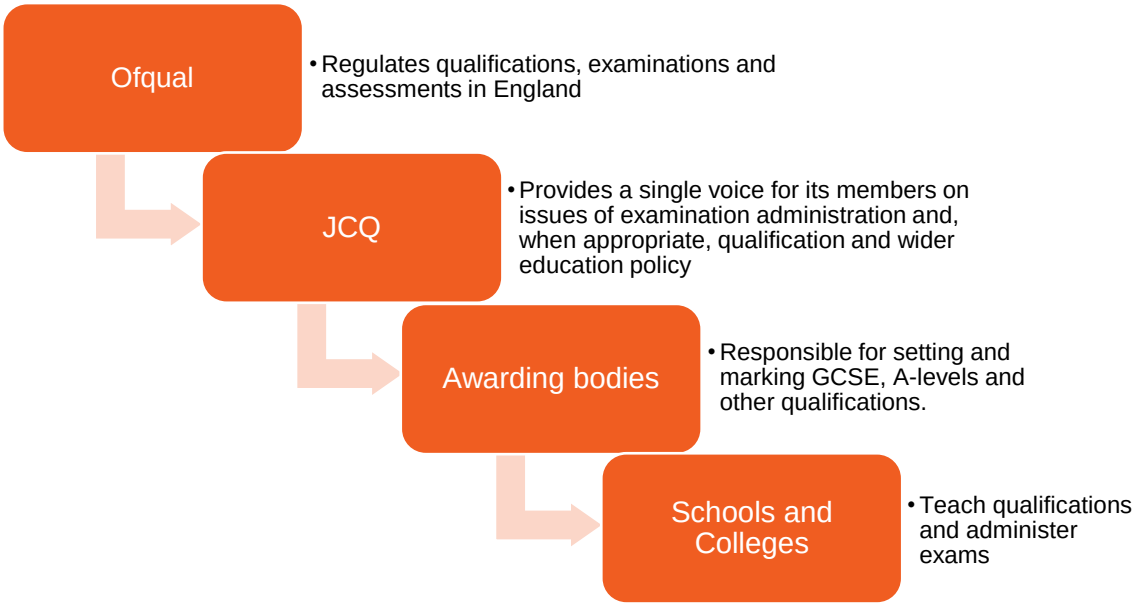
[13] GCSE raw marks are less comparable across years than grades. A workaround is to calculate decimal grades, specifying at a more precise level where learners' scores sit within standardised grade boundaries. For example, for a score one-tenth of the way through the mark range at grade 5 in a GCSE, the decimal grade would be 5.1, 20 per cent of the way through the grade would be 5.2, and so on.

of learners sitting the exam with each board. As such, in addition to accessing GCSE marks, it is necessary to obtain these parameters from awarding bodies.

For an even more sensitive measure of an intervention's impact, evaluators can assess GCSE papers to determine which questions are most related to the intervention, and create a 'subscale' using item-level data. This approach was used in the evaluation of the CfEM's Teaching for Mastery (TfM) programme, and the evaluators of the Deeper Thinking pilot conducted a similar exercise to determine what a plausible impact of the intervention might be on pupils' GCSE Science grades (EEF 2020).

While GCSE grades are available in administrative sources, obtaining raw marks and the population parameters needed to convert these into standardised scores is administratively complex. GCSE *grades* are stored in the NPD[14], which is typically used in EEF evaluations, as well as the ILR (GCSE English and Maths only). On the other hand, GCSE marks are only held by secondary schools (Key Stage 4 assessment) and post-16 settings (GCSE resits), as well as awarding bodies (to whom papers are sent for marking, developing grade boundaries and converting to grades) and Ofqual[15] (for regulatory purposes).

Figure 1. Examination governance



Evaluators on recent EEF-funded projects have chosen to access GCSE marks from the settings that learners are in when resitting the exam, and to access prior attainment data from the NPD for use as a baseline measure. The baseline measure typically used is Key Stage 2 scores. Sometimes, the

[14] GCSE grades are available in the Key Stage 4 performance table of the NPD, regardless of when learners are sitting the exams. In the case of resits, Key Stage 4 data is replaced with the most recent grade. Nonetheless, there is a table which tracks historical performance data.

[15] The only administrative data source containing GCSE marks is the GRading and Admissions Data for England (GRADE), which was established in the context of the Covid pandemic and the alternative awarding arrangements put in place for progressing to post-16 education when exams were cancelled in 2020. GRADE is a data sharing project betweenOfqual (data on GCSE and A level examinations and qualifications collected from awarding organisations), the DfE (extracts of the NPD for GCSE and A level learners) and UCAS (data from the university application process) covering exam results from 2017 to 2022. It aims to rebuild trust and confidence in the assessment system by allowing external researchers to conduct independent high-quality evaluation of the judgements made in awarding grades in 2020 and, more broadly, to enhance the quality of the assessment system and produce evidence to inform future education policy.

previous GCSE *grade* has been used as an additional covariate, but it does not offer very much explanatory power[16].

Accessing GCSE marks data from settings is very resource intensive, requiring a high level of cooperation from various actors within a setting (e.g., principals, data officers, teachers, etc). It (and data collection more generally) has been the most challenging aspect of EEF-funded post-16 evaluations to date, often resulting in high levels of missing data. Nonetheless, it is the only 'tried and true' method for obtaining the GCSE raw marks[17] and there are steps that can be taken to improve the completeness of data returns from settings. These include:

- Having a dedicated member of the evaluation team responsible for liaising with settings (and planning for this to be a full-time job during data collection periods).
- Identifying the key contacts in settings and establishing preferred communication strategies. For collecting GCSE marks, the key contact is typically the 'data officer' or 'MIS manager' in the college. However, input from teachers or other staff may be necessary, and having multiple key contacts for each setting is recommended to safeguard against staff turnover, sick leave, etc.
- Factoring in enough time to build good working relationships with settings and maintaining regular contact throughout the trial period (especially for control settings).
- Planning to collect GCSE candidate numbers[18] and/or Unique Candidate Identifiers (UCI)[19] from settings, alongside learners' ULN, to be able to match GCSE marks data with the NPD. Note that exam boards primarily use these two numbers rather than the ULN to identify learners. Other helpful data to collect includes exam board and centre number[20].
- Providing guidance to prepare settings for sharing data (e.g., a data collection handbook including flow charts and timelines) and creating templates for data returns that minimise the amount of prep required.
- Monitoring the quality of data as it is received from settings, building data validation and cleaning into the process early on.
- Getting the timing right for requesting the data and allowing plenty of time for chasing settings. Note that exam results come out in August, which is when data officers will be able to download GCSE marks data from their exam board's web portal. However, a few months of chasing may be required (e.g., until October).

Secondary outcome measures and other data collection

In addition to the principal attainment outcome, post-16 evaluations are often concerned with secondary, more proximal outcomes[21]. These outcomes are analysed either in the impact evaluation or implementation and process evaluation (IPE), to validate areas of the theory of change. At the learner

[16] This is because most GCSE resit students will have achieved a grade 3 on their first attempt; those with a grade 1-2 are often entered into a FSQ Level 2 exam instead of a GCSE resit. As such, there will be little variation and explanatory power from GCSE grades alone.

[17] The EEF is currently working with others to facilitate access to GCSE marks data in administrative sources. This includes conversations with Ofqual to better understand how the GRADE data sharing project can be leveraged for use in EEF-funded post-16 evaluations. It also includes conversations with the DfE to explore the feasibility of having GCSE marks data added to the NPD. Another alternative under consideration is to access GCSE marks from awarding bodies, either individually or through the JCQ.

[18] This is a four-digit number generated by the school or post-16 setting in which the learner is sitting the GCSE exam (centre). It is unique for the exam series, meaning learners may have different numbers each time they sit the GCSE.

[19] This is a 13-character code generated by the school in which the learner first sat the GCSE exam (centre), according to JCQ specifications. It is unique for each learner and used to link their results across exam series (and across different centres).

[20] This is a five-digit number generated by JCQ for the school or post-16 setting in which the learner is sitting the GCSE exam.

[21] Given their policy relevance, GCSE *grades* are also always included as secondary outcomes in EEF post-16 evaluations. The measure can be the actual grade, the probability of achieving a grade 4 or the probability of moving up a grade.

level, these outcomes may include attendance, and other intermediate outcomes which might mediate the relationship between receiving an intervention and improved attainment outcomes.

Attendance

Interventions may seek to improve attendance and engagement with GCSE resit classes on the path to improving GCSE attainment. Even if they do not aim to increase attendance rates, an intervention's impact often depends on learners attending their GCSE resit lessons and/or engaging with the key inputs of the intervention. As such, evaluation designs may often include assessing the impact of the intervention on attendance or including attendance in analyses of compliance and/or dosage. However, as there is no system in place for monitoring attendance at a regional or national level, the only way to access this data is from post-16 settings themselves. As with GCSE data, it is important to consider the best approach to ensure the completeness of this data, such as establishing clear expectations and lines of communication between the evaluator and settings. Furthermore, evaluators will need to consider the comparability of the data across post-16 settings, as there is likely to be variation in how they record attendance (e.g., measuring attendance as a proportion of all possible resit lessons), as well as the accuracy and precision of the data (e.g., data returns for the entire academic year vs the intervention period) and the likelihood of skewed distributions.

Other proximal outcomes

The GCSE resits review highlights that, aside from subject knowledge, other skills and factors may be important to achieving a grade 4 on the GCSEs, particularly for learners who have not achieved a grade 4 on several attempts. These may include confidence, self-efficacy, resilience, motivation, perceptions about the importance of doing well on the GCSEs, etc. There are many appropriate measures available for these secondary outcomes, and the SPECTRUM database is a good starting point. While some of these measures have been successfully administered in EEF-funded evaluations at KS3, the post-16 context may make the administration of assessments and questionnaires more challenging (especially at baseline).

Collecting secondary outcome and IPE data from learners and settings

It is important to carefully consider the timing for baseline and endline data collection, and to coordinate the administration of the measures with settings. As mentioned earlier, learners may only select their setting and their courses in October. In addition, variable attendance rates should be factored into the data collection process when collecting data from learners directly (e.g. asking learners to complete a questionnaire), and evaluators may need to provide multiple opportunities to collect data from learners to avoid high rates of missing data. It is important to ensure that secondary data collection and IPE activities are not overly burdensome to learners and staff, and do not excessively disrupt teaching/intervention activities.

A potential alternative to administering measures in settings may be reaching out to learners directly with SMS surveys. As learners provide their own consent in post-16 research, it should be feasible to send surveys directly to learners, provided that this comes with adequate information about their data protection rights and allowing them the opportunity to withdraw. For instance, the EEF's Texting Students and Study Supporters trial delivered the intervention and collected data through SMS messages to students[22].

[22] College tutors introduced learners to the trial in their classrooms using materials provided by the developer, the Behavioural Insights Team (BIT). In-class time required for the trial was minimal, only requiring learners to complete a 10-minute online form in which they nominated a study supporter via BIT's Promptable website during the first weeks of term. In this form, learners registered their interest, nominated a study supporter, and answered some questions about their learning behaviours. Note that participation was voluntary for learners, and they indicated their consent via the Promptable form.

Recruiting for and designing a well-powered study

A key criterion in EEF-funded trials is statistical power, and one of the main determinants of this is sample size – having enough trial participants. While the recruitment and retention of participants to a trial is always a challenge, it is even more so in the post-16 sector because there are fewer settings to recruit from and the variation in how they are structured makes coordination more difficult. Furthermore, since the sector has had fewer opportunities to engage in rigorous research in the past, understanding of research methodologies such as randomised control trials (RCTs) and interest in engaging with research may be more limited. The EEF is currently developing its support offer for delivery and evaluation teams recruiting settings to post-16 evaluations.

Recruitment and retention of practitioners and learners may also be challenging. The GCSE resits review found indications that some practitioners and learners question the value of GCSE resits relative to overall post-16 provision. On the one hand, practitioners who are subject specialists may prefer to teach A-levels over GCSEs or perceive function skills curricula as more suitable for vocational learners despite their setting's policy being to enrol learners in GCSEs. On the other hand, as resit learners are often studying technical or vocational qualifications, they may not always have positive attitudes towards reengaging with GCSE courses. As such, evaluators may need to explore in more detail the attitudes of the teachers and learners taking part in the evaluation. Additionally, there is a smaller pool of GCSE resit teachers and learners to recruit from for evaluations in the post-16 sector (e.g., practitioners often teach multiple resit classes and resit learners are only a subset of all 16-19-year-old learners) and learners' variable attendance patterns make accessing and keeping in touch with them difficult.

To increase the power of post-16 evaluations to detect an effect on GCSE attainment, given these issues with recruitment and retention, study designs may look different to other education contexts (e.g., the school sector). A key question here is selecting the appropriate level of randomisation. Below are some potential levels of randomisation, starting from the highest cluster to the lowest individual level.

- **Setting:** Randomly assigning settings into a treatment/control condition, such that all campuses within the same college would be in the same condition as each other.
- **Campus:** Randomly assigning campuses into a treatment/control condition, such that different campuses within the same college may be in different conditions.
- **Teacher:** Randomly assigning teachers to a treatment/control condition.
- **Class:** Randomly assigning GCSE resit classes to treatment/control conditions.
- **Learner**: Randomly assigning individual learners to treatment/control conditions.

Most school-based EEF-funded trials have followed a cluster randomised design, randomising at the school level to avoid spillover (whereby pupils in the control group receive the intervention when they should not, or vice versa), though some have randomised at other levels (e.g. pupil- or class-level). However, post-16 settings, especially FE colleges, are less numerous, and typically larger, than schools. Moreover, it is common for FE colleges to have multiple campuses, and campuses within colleges can be geographically distant from one another and differ substantially in their curricula and learner profiles. Consequently, it is important to consider whether FE colleges with multiple campuses should be considered a single cluster for randomisation or treated separately, with the latter often being most appropriate[23].

As can be inferred from Table 1, the number of GCSE resit learners varies between settings, with FE colleges serving a larger proportion of this cohort than school sixth forms. As substantial variation in the

[23] Analogous considerations exist in other EEF trials. In school-based trials, randomisation is typically clustered at the school level, but careful consideration is given to whether to include multiple schools from a single multi academy trust (MAT) in the same evaluation. A similar approach is taken in the treatment of nursery groups in Early Years trials.

size of clusters may reduce statistical power, evaluators should consider whether to focus the evaluation on a single setting type (which may depend on the programme or intervention) or account for cluster size variation in their sample size calculations. The latter is often done by incorporating an inflation factor known as the coefficient of variation into power calculations.

For some programmes, where spillover effects are unlikely to be large, teacher or learner level randomisation may be most appropriate. For instance, individual tutoring programmes may be suitable for individual randomisation at the learner level, particularly if the tutoring is delivered one-to-one. Meanwhile, professional development programmes could potentially be randomised at the teacher level. Note that the GCSE resits review highlighted there is little collaboration among teachers of resit classes in the post-16 sector, especially in FE colleges. Class-level randomisation is another option, although this may be logistically challenging as teachers often teach multiple resit classes and, depending on the intervention, it may be hard for them to switch from 'intervention' to 'business as usual' between classes.

Evaluators, in dialogue with developers and the EEF, should select the most appropriate level of randomisation for the specific trial in question and adequately account for it in their power calculations. Parameters used for power calculations, and the assumptions they are based on, should be backed up by clear evidence/rationale. For reference, we include a summary of the parameters used for power calculations in post-16 trials to date in Appendix B.

Practical issues in implementing an evaluation design

After a trial has been designed, with a randomisation and sampling strategy, there remain a number of practical challenges to implementing the design in practice. Three key considerations are (1) enumerating learners, baselining, and randomisation, (2) tackling attrition and non-response, and (3) deciding how to handle November resits.

Enumerating learners, baselining, and randomisation

In school-based trials, the EEF generally recommends enumerating the pupil sample and conducting baseline tests (if applicable) before randomisation to minimise bias. This can either be done at the start of the academic year in which the trial will be implemented, or the previous academic term before schools break for the summer holidays. At post-16, learner lists are not finalised until mid-to-late October, so it may not be feasible to enumerate learners until the start of November. If this is the case, then evaluators should carefully consider the risk of bias. For example, group allocation can be concealed from learners (e.g., by asking teachers to withhold this information) until after enumeration is complete. As for baseline testing, EEF-funded post-16 trials to date have relied on KS2 data from the NPD (for the primary outcome, and in some cases for secondary outcomes) to minimise burden. Evaluators who are considering baseline testing for secondary outcomes would need to weigh the pros and cons of doing so.

Attrition and non-response

It is also important to consider the implications of the recruitment, retention and data collection challenges mentioned above on the security of findings, and carefully consider attrition and non-response when designing and implementing a post-16 evaluation. Taking the example of an individually randomised tutoring programme, there are two main points at which withdrawals from both the intervention and evaluation might occur, each bringing related but distinct challenges.

First, learner drop-out may occur after enumeration, baselining, and randomisation, but before the intervention delivery commences. Anecdotally, this is a prevalent issue in the post-16 sector, with settings accustomed to learners signing up for support and then not attending. This effectively means that there are learners within the treated group who did not receive the intervention at all, or who receive the intervention but withdraw from the evaluation. Furthermore, the challenges of collecting data from

settings may mean it is not possible to enumerate an adequate number of learners in the first place and/or collect primary outcome data for them at endline. Because the headline finding reported in EEF trials is intention-to-treat (analysing learner data as they were randomised, not based on actually receiving the treatment), and it is not possible to include learners who have withdrawn from the evaluation/for whom data is unavailable, this can dilute impact estimates[24].

Second, learner drop-out may occur further into the programme (e.g., after 5/10 tutoring sessions) or at the point of collecting endline data. In the former case, learners can more reasonably be viewed as having received the programme of interest. Indeed, attrition at this point in the programme can be indicative of limitations of the programme itself. However, this may be difficult to explore if learners cease to engage with the *evaluation* or if it is not possible to collect endline data for them.

Attrition and non-response also occurs at the setting and teacher level, for instance due to staff turnover, sickness, or workload issues. Overall, the following steps can be taken to mitigate the impact of attrition:

- Offer extended induction to programmes, to enhance participant investment early in the trial process.
- Carry out an in-depth implementation and process evaluation which explores attrition at all levels and timepoints, identifying causes and potential solutions.
- Communicate to settings and developers the importance of preserving randomisation integrity and not offering the intervention to control group participants to fill slots left by intervention group attrition.
- Explore options for the timing of enumeration, randomisation, and baselining to minimise post-randomisation attrition.
- Minimise data collection burden for learners, teachers, and settings.

November resits

Furthermore, the fact that learners may resit their GCSEs in the autumn term (November) rather than waiting for the following summer makes analysis more complicated. November resits raise concerns for fidelity and dosage, as learners enrolled for the November resit may cease to attend classes once they have taken the exam, even though results are not released until the beginning of the next calendar year (January).

In some cases, it may be preferable to exclude learners taking November resits from evaluations to ensure comparability and consistency within an evaluation. However, it may be the case that a more intensive revision programme could prepare some learners for November resits, and so the exclusion of these learners should be taken with caution. The best approach will depend on the programme being evaluated. For instance, CfEM's TfM programme was designed to be delivered across a full academic year, with effects not expected until the summer resits, and so November resit learners were excluded from the evaluation. Alternatively, the 5Rs trial explored differential impacts for November and Summer resit learners, while the BMP evaluation included November resits in their primary outcome measure of the probability of achieving a grade 4 in GCSE maths over the course of the pilot.

Defining business as usual

The practice review finds little evidence of holistic thinking or sector-wide approaches in relation to approaches for GCSE revision. That is, everyone is technically doing the same thing in these resit classes, but there is no standardised way of doing it, with substantial variation between settings. Moreover, it is common for FE colleges to have multiple campuses, and campuses within settings can be geographically distant from one another and differ substantially in their curricula and learner profiles.

[24] EEF trials often perform Complier Average Causal Effect (CACE) as a secondary analysis, which looks at the impact of the programme on those who actually received the intervention (which is operationalised variously) but compromises the integrity of the randomisation.

As such, defining business as usual is difficult because of the range of different practices being implemented and potential overlap with a particular intervention.

EEF-funded programmes are selected on the basis that they are expected to be distinct from 'business as usual', but it is important for evaluators working in the context of multiple concurrent interventions to consider whether settings, especially those in the control group, are likely to be doing something similar to the programme being evaluated. It is thus essential to develop a comprehensive understanding of the programme, its core components and causal mechanisms, to carefully consider eligibility criteria for the evaluation, and to carry out a robust implementation and process evaluation which monitors activities within the control group (i.e., under 'business as usual' conditions).

Selecting an appropriate indicator for disadvantage

The EEF is dedicated to breaking the link between income and educational achievement, and all EEF-funded programmes should contribute to reducing the socioeconomic attainment gap. As such, EEF asks evaluators to include subgroup analysis, split by an indicator of socioeconomic disadvantage[25].

The EEF's strongly preferred indicator for disadvantage for evaluations is FSM status, specifically an indicator of whether a learner has in the last six years been eligible for free school meals (EVERFSM_6_P). This indicator is available in the NPD until learners reach Year 11 and is one of the indicators used to designate eligibility for Pupil Premium.

Although there is no Pupil Premium in the post-16 sector, free meals are in fact available to post-16 learners and settings report this eligibility in the school census/ILR. However, because the process of applying for FE free meals and verifying eligibility is cumbersome, take up is likely to be lower in post-16 settings than schools. As such, this is likely not a very accurate indicator of socio-economic disadvantage.

While the EEF is currently working to determine the best indicator for socio-economic disadvantage in post-16 research, current guidance is to use the 'historical indicator' of EVERFSM_6_P from Year 11.

Supporting engagement with EEF-funded research

Compared to the school sector, in which the EEF has been funding large-scale trials for over 10 years, exposure to/knowledge of RCTs is limited is in the post-16 sector. For instance, there is relatively limited understanding of the importance of randomisation and having a control group when trialling programmes. Settings may have ethical concerns about randomisation, and it is important to address these while communicating the benefits of taking part in trials for settings and learners in the short term, and for the wider evidence base and sector as a whole in the long term.

While the EEF is currently working to improve its engagement with the post-16 sector at a higher level, evaluators should be prepared to support delivery teams with communicating key principles and aspects of impact evaluation design (e.g., randomisation, statistical power, control groups) to settings at a project level. It is important to keep in mind that post-16 setting structures are quite different from schools, and there is substantial variation between and within the different types of post-16 settings. Moreover, the structure of the workforce at post-16 significantly differs from other educational stages (e.g., more variation in contract type, qualifications, and professional background). As such, different approaches for supporting engagement with EEF-funded research will likely be necessary among settings and practitioners.

[25] Subgroup analysis is typically carried out either using an interaction term of the indicators for treatment and FSM status, or by re-running analysis for FSM and non-FSM learners separately.

For EEF-funded trials, it is common to pay a financial incentive to control settings (and in some cases, to individuals for any evaluation-related activities they engage in) to acknowledge their contribution to the research. However, in the context of post-16 settings, the amounts we are able to offer to settings may be small compared to their overall budget. Nonetheless, it might be useful to reframe setting-level incentives, for instance framing them as the cost of the time it would take the data manager to provide the data for the evaluation. Beyond incentives for participating in research, it might prove necessary to ensure staff cover so that teachers can engage with the *intervention* and any prerequisites (e.g., training, planning time). Given high teacher workloads and learners' variable attendance patterns it may be appropriate to provide incentives directly to staff/learners rather than to settings.

Besides incentives, non-financial support to settings may also facilitate higher rates of data returns, for instance by sending researchers to the setting to support with data collection activities. Given structural differences among post-16 settings, it may help to adjust which members of staff are contacted for different evaluation-related activities. For instance, while principals are ultimately required to 'sign off' evaluation-related engagement, it may be more appropriate to liaise with other figures (department/subject leads, teachers) for some circumstances and purposes.

Note that, despite their limited engagement with RCTs to date, settings do engage in research more generally, and often encourage practitioners to run action research projects to improve their practice. Evaluators should be aware of any teacher-led research that may be taking place alongside their evaluations.

IV. Conclusion

The EEF has faced several challenges in the research it has commissioned to date to support GCSE resits. Nonetheless, the evidence gaps for the post-16 phase of education, and how important it is for supporting socioeconomically disadvantaged learners before they leave the education system, means there are significant benefits for educators and policy makers if we can tackle them. To that end, this paper highlights some early considerations and recommendations for evaluations in the sector, which we plan to update regularly based on feedback received and our continued and expanded work in the post-16 sector. We would be very grateful for any feedback on this paper and any advice on increasing the success of post-16 interventions and evaluations. Please send your feedback to info@eefoundation.org.uk, marking your email for the attention of the Evaluation team.

Appendix A: Summary of the EEF's post-16 evaluations

To date, the EEF has funded five evaluations of post-16 programmes aimed at supporting GCSE resits[26], which have been either pilots or efficacy trials, as well as an impact evaluation of a policy pilot using a quasi-experimental design (QED). There are several challenges to running post-16 evaluations, and the EEF has commissioned research to explore some of these[27]. Between 2020 and 2021, Alpha*Plus* and Manchester Metropolitan University were commissioned to explore the difference between using GCSE grades and raw marks as a primary outcome and develop guidance around choosing between the two. More recently, the Centre for Education & Youth and the University of Warwick were commissioned to complete a practice review to build a more robust and objective picture of what current practice looks like for GCSE resits. While this was primarily done to inform the EEF's funding of its programme pipeline, it also provided insight into 'business as usual' in the sector.

Table 1: EEF-funded post-16 evaluations

Evaluation	Type	Subject area	Primary outcome	Results
5Rs approach to GCSE Maths resits (5Rs)	Efficacy trial	Maths	GCSE raw mark	The first efficacy trial was affected by the Covid pandemic, resulting in only IPE results being published. Due to recruitment and data collection issues, the retrial was adjusted to a split cohort design and the second cohort eventually cancelled, meaning that only IPE results will be published.
Maths-for-Life	Efficacy trial (including a formative pilot)	Maths	GCSE raw mark	The publication of the evaluation report for this trial has been delayed due to data access issues.
Texting Students and Study Supporters (Project Success)	Efficacy trial	English / maths	GCSE grade	This trial did not find evidence that the intervention had any impact on FE learners' rate of 'passing' their GCSE resits (or on improving their attendance).
Embedding contextualisation in English and mathematics GCSE teaching	Pilot	English / maths	-	This pilot found limited increases in the use of contextualised learning in the classroom. It was therefore difficult to assess whether the intervention had an impact on outcomes like retention and attainment.
Assess for Success	Pilot	English	-	This pilot found some evidence of promise, but recommended significant development and further piloting before the programme would be ready for a further trial.
Basic Maths Premium Pilot (BMP)	Policy evaluation (QED)	Maths	GCSE grade	This policy evaluation did not find evidence that additional funding had any impact on improving resit learners' level 2 maths attainment (or on the likelihood that eligible learners sit a GCSE rather than a Functional Skills Level 2 exam/ no resit exam at all).

[26] The EEF has recently commissioned its first post-16 effectiveness trial, building on the findings of the evaluation of the CfEM's Teaching for Mastery programme, as well as a pilot of a targeted tutoring programme for disadvantaged GCSE English and Maths resit learners. We are currently commissioning two additional programme evaluations (one pilot and one efficacy trial) in the post-16 sector.

[27] Between 2015 and 2016, the EEF also commissioned a literature review on improving Level 2 English and maths outcomes for 16 to 18 year olds. More synthesis work is being planned in the post-16 sector.

Appendix B: Summary of the parameters used for power calculations in post-16 trials to date

Parameter	Explanation
Level of randomisation	Whether randomisation is done at the individual level or clustered, and if so at which level.
Sample size	The number of learners, settings, and campuses in each treatment arm (and by subject, if the evaluation includes both English and maths outcomes). If clustering, a further parameter to consider is the number participants per cluster (mean cluster size).
Coefficient of variation	A standardised measure of dispersion on outcome scores, which is typically included to account for variation in cluster sizes. In the 5Rs approach to GCSE Maths resits trial, a coefficient of variation of 0.33 was used, based on the evaluators' calculations.
Pre-test/post-test correlation	The correlation between outcome scores before and after the intervention. Assumptions in post-16 trials to date range from 0.5-0.6, which are conservative estimates based on correlations between KS2 and KS4 maths examinations that are revised down to assume the correlation will be lower for resit learners (Maths-for-Life), KS2 and GCSE maths examinations (5Rs). Some evaluators were even more conservative and did not adjust for variance explained by including a baseline measure (BMP, Project Success, CfEM TfM)
Intra-cluster correlations (ICCs)	The relatedness of outcome scores within clusters. Assumptions in post-16 trials to date range from 0.1-0.27. Typically based on earlier trials (5Rs retrial, Maths-for-Life, CfEM TfM). BMP used the highest estimate (0.27) based on similarity in eligible learners pre-intervention.
Significance level	The likelihood of falsely rejecting the null hypothesis (false positive). Typically set to 5%.
Statistical power	The likelihood that a test will detect an effect of the desired size if there is one. Typically set to 80%.
One-sided/two-sided hypothesis	Whether the hypothesis being tested is directional (one-sided) or non-directional (two-sided). Typically two-sided, which is more conservative.
Attrition	Estimated loss to follow-up. Estimates vary and will differ between programmes. Assumptions for post-16 trials to date range from 15-25%.
Baseline GCSE resit rate	The proportion of learners attempting GCSE resits at baseline. Assumed to be 30% in 5Rs retrial, based on previous trial.

Baseline GCSE 'pass' rate	The proportion or learners passing their GCSE resits at baseline. Assumptions for post-16 trials to date range from 25-28%, though calculations based on data from the previous two years suggest the 'pass' rate for resit learners range from 13-23%.
Desired minimum detectable effect size (MDES)	The attainable MDES will be affected by the intervention type and the outcome measure (i.e.., GCSE marks vs. grades vs. decimal grades).
Other assumptions	When using a **simulation approach** for power calculations (as was done in Maths-for-Life), additional assumptions include: • The **distribution of GCSE scores** follows a normal distribution (based on Ofqual data). • The simulated **treatment effect is uniform** across all participants in the treatment group (based on lack of evidence to the contrary).

Table 2. Parameters for power calculations in post-16 trials

Project	Evaluation type	Intervention	Outcome	Level of randomisation	Sample size	Mean cluster size	Pre-post test correlation	ICC	MDES and/or effect size achieved	Attrition
5Rs (retrial (Covid-interrupted)	Efficacy	Whole class	GCSE Maths score	Setting	88 (design)	51 (design)	0.6	0.17 (design)	0.23	Setting level: Not reported Learner level: 15% (assumption)
5Rs (retrial)	Efficacy	Whole class	GCSE Maths score	Setting	80 (design)	50 (design)	0.6	0.17 (design)	0.23	Setting level: Not reported Learner level: 15% (assumption)
Maths-for-Life	Efficacy (including formative pilot)	Whole class	GCSE Maths score	Setting	50 (design)	73 for FE colleges; 32 for other settings (design)	0.5	0.2 (design)	0.23	Setting level: Not reported Learner level: 20% (assumption)
Basic Maths Premium	QED	Whole setting	GCSE Maths grade	Setting	679 (analysis)	4.63 (analysis)	No adjustment	0.20 (design) 0.28 analysis)	1.36 (odds ratio) 1.008 [0.834, 1.214]	Setting level: 51% Learner level: Not applicable
Project Success	Efficacy	Targeted	GCSE English/ Maths grade	Learner	31 (analysis)	59 (analysis)	No adjustment	0.10 (design) 0.28 (analysis)	6.4pp 0.01 [-0.05; 0.07]	Setting level: Not applicable Learner level: 0.9%
CfEM TfM (Covid-interrupted)	Efficacy (not funded by the EEF)	Whole class	GCSE Maths score	Setting	112 (analysis)	21 (analysis)	0.25 (learner level) 0.23 (setting level)	0.20 (design) 0.14 (analysis)	0.25 0.06 [-0.12, 0.24]	Setting-level: 23.8% Learner level: 54.5%