

Durham University
Dr Rebecca Morris, Prof. Stephen Gorard
Prof. Beng Huat See and Dr Nadia Siddiqui

Evaluation Summary	
Age range	Secondary – Key Stage 4 (Years 10 and 11)
Number of pupils	c. 12,500
Number of schools	100
Design	Randomised controlled trial with randomisation at school level
Primary Outcome	English Language and English Literature GCSEs (cancelled)
Secondary outcome	Teacher workload (related to marking and assessment)
Protocol date	February 2021
Version	3
Changes to the protocol	- Updates to the protocol and SAP following the cancellation of externally-awarded GCSEs in 2020 and 2021 as a result of the COVID-19 pandemic. Thus resulting in no primary outcome for the evaluation. This updated protocol was written after the delivery of the intervention (2018-2020) and prior to the collation of the final evaluation report.

Evaluation Protocol for Independent Evaluation of ‘FLASH Marking’

Background

The concept of providing high-quality formative feedback which encourages students to identify and develop their strengths and areas of improvement has had considerable attention in relation to research, policy and practice (see e.g. Black and Wiliam, 1998; Christodoulou, 2017; Hattie and Timperley, 2007). To date, however, there have been few large-scale, robust studies that have examined the use of written feedback approaches on pupils’ work (Elliott et al., 2016).

A number of studies cited here have explored the use of grades and/or comment-based feedback on students’ written work. The evidence seems to suggest that grade-only marking does not have a positive impact on attainment (Elliot et al., 2016). Comment-only feedback (with no grades) was also found to be less beneficial when compared with both a comments and grades approach (Gorard and Smith, 2005). Another study, however, found that detailed, descriptive feedback was more effective for students’ progress when given alone, unaccompanied by grades (Lipnevich and Smith, 2009).

The use of a combination approach (of comments and grades) has produced less clear research results. Work by Klapp (2015) showed that boys and lower attaining pupils did less well when receiving grades and comments whereas there was a positive long-term effect for girls. This study, based in Sweden, using data

from 30 years ago, focused on the feedback students received in primary school and their latter attainment in secondary school. While the findings are important, it is not clear how applicable the findings are to a different context and a shorter term approach for understanding the impact of grades/comments on academic outcomes. A smaller scale study of middle school science achievement in America (Zhang and Misiak, 2015) found that a combination of both written comments and grades was most effective for promoting student attainment. The authors concluded, however, that in order to move students' learning forward, the comments must be highly relevant to the academic standards being assessed.

In addition to considering the benefits of different forms of formative assessment, it is also important to consider the costs too. Concerns surrounding the workload implications of extensive Assessment for Learning (AfL) practices combined with very little evidence that they work have led to calls for a pared back approach, particularly in relation to the amount of time that teachers must spend marking written work (Independent Teacher Workload Review Group, 2016). There is a renewed interest, supported by government, Ofsted and the unions, in finding more efficient and effective methods for providing valuable feedback. While this has led to some interesting and innovative approaches, more rigorous testing of these is needed in order to assess their effectiveness and scalability.

It is also important to situate this evaluation within the current policy context regarding assessment and schools. The removal of National Curriculum levels has provided additional autonomy to schools, allowing them to develop their own methods for supporting and measuring attainment and progress (Gibb, 2015). Many in the profession have welcomed this removal of a prescriptive approach but there is also an acknowledgement that many schools are not clear about the best approaches for replacing levels. The government have suggested a range of alternative methods (Lilly et al., 2014) although there is often a lack of rigorous evidence underpinning these. Changes to the English GCSE grading system and the exam specifications (in relation to content and a shift to a linear rather than modular model) (Ofqual, 2015) mean that many schools are looking for improved ways to develop students' knowledge and skills in preparation for the exams and later life. The use of good quality formative assessment of written work may be an important element in supporting this.

While the analysis of pupils' GCSE outcomes (the primary measure for the evaluation) is not possible, this project will still produce a final report outlining school, teacher and pupil engagement with the FLASH Marking intervention, as well as drawing upon our analysis of the pre/post teacher workload survey. It is intended that this report will provide important information and insights for school leaders and practitioners who may be thinking about feedback policy and practice in their settings, and may also help to inform future evaluation work in this area.

Intervention

FLASH (Fast Logical Aspirational Student Help) marking is a school-developed feedback approach in which teachers give skills-based comments rather than grades in Key Stage 4 English. The approach was devised by staff at Meols Cop High School in Southport. The FLASH feedback uses language from the highest grades descriptors in GCSE English Language and English Literature. They are presented on students' written work using a code system which identifies where pupils have done something well but also where there are areas for improvement and development.

The codes are broken down into sections depending on whether they are describing skills focusing on reading or writing skills. During the development stages, they have been used by both teachers and pupils in English lessons. They are used to inform target-setting for future pieces of work.

The developers outlined a number of reasons for devising and implementing FLASH marking. First, it was hoped that in having a clear skills-focus during teacher, peer and self-assessment that students would be able to clearly understand how they can improve and develop in English. Second, by removing the use of grades, it was hoped that pupils will focus on the skills and knowledge that they need to develop in order to improve their attainment. Third, the use of codes is believed to 'speed- up' the marking process and ensure that

children receive feedback sooner. Closely linked to this was the aim of reducing teacher workload in relation to assessment. Finally, tracking of pupils and classes through the coding system can highlight where there are skills/knowledge gaps and these can be addressed. The intervention was piloted in a small number of schools prior to full roll-out for the purposes of the efficacy evaluation.

Two English teaching representatives from each 'intervention' school attended three training sessions (run by the Meols Cop development team). These staff were then responsible for cascading the training to all other staff within their department. The school subject leader for English was one of the staff members responsible for attending the training and they were asked to select one other English teacher to attend too.

Details of the three training sessions are below:

- Training session 1 ran for a full day in June 2018. It focused on the principles of FLASH marking, lesson planning, modelling the assessment approach, using demonstration videos and the use of the web portal as a support mechanism for trial participants.
- Training session 2 was a three-hour session in November 2018. This session focused on moderation of assessments and the use of the codes in future planning. Support was offered through the use of demonstration videos and group discussion.
- Training session 3 ran in September 2019, as schools enter the second year of the trial. This session ran for approximately four hours and included two main elements: the first part was used as 'refresher' session for schools that needed this or who had new staff who need initial training. The second part was for two nominated English staff from participating schools and provided an opportunity to discuss progress and receive further support from the delivery team.

Delivery of the training took place in six geographical 'hubs'. This allowed groups of schools from the same area/region to come together and receive the training as a group.

Additional support was made available to English departments throughout the period of the trial. This was provided by the development team throughout and included: frequent contact between schools and the development team; visits from the development team if needed; and the use of the web portal as a way of providing videos, models of assessed work and curriculum resources. It also involved communication between the subject leader and development team, the whole English department or specific teaching staff, depending on what the school's needs were.

Research Plan

This outline reflects the revised research plan following the government's announcement to cancel externally-awarded GCSEs in 2021.

Research questions

The primary research question for this evaluation was:

- How effective is FLASH marking in improving the GCSE English outcomes for Key Stage 4 pupils?

Without the GCSE data, it is no longer possible to respond to this question.

The evaluation also included a second research question. Using the pre/post teacher workload data, we will provide a response to this question as originally planned.

- How effective is FLASH marking in reducing the marking and assessment workload for teachers of Key Stage 4 English?

Design

A two-arm RCT with randomisation occurring at school level was run between 2018-2020. Following the cancellation of externally-awarded GCSEs in 2020, the evaluation team and the EEF explored the possibility of an extension to the evaluation in order to respond to the primary research question on attainment.

However, with the cancellation of the 2021 external GCSEs too, this was no longer a viable option and decision was made to abandon plans to evaluate the academic effectiveness of the intervention using attainment data. Below we outline details of how the project and evaluation was set-up in order to provide some context to the final analyses that were able to occur (based upon workload survey data and process evaluation data). For further details about the original trial, we recommend that readers refer to earlier versions of the protocol published on EEF's website¹.

A total of 103 schools participated in the FLASH Marking evaluation. In Spring 2018, 52 of these were randomised to the intervention group and 51 were randomised to the control group. Those in the 'intervention' group received the FLASH marking programme, including the training and cascading of training and implemented the approach with their Year 10 classes from September 2018; the intervention continued with these same groups until the end of their GCSE course in Year 11 (June 2020). Those in the 'control' group operated a 'business as usual' approach, continuing with their usual methods of marking and feedback.

Randomisation

Schools will only be eligible for randomisation after:

- Signing a Memorandum of Understanding
- Providing pre-test data requested in the Memorandum of Understanding (including pupil UPNs, prior attainment and demographic data and providing teacher contact details).
- Completion of the first staff workload survey. All English department staff will be asked to complete this short online survey.

Randomisation will be conducted at the school level once all of the above data has been collected. The process of randomising schools to each group will be carried out independently by the evaluation team at Durham University, and observed by at least one colleague. Each school will be assigned a random number, created in Excel, and the list of schools will then be sorted into number order. The first half will be the treatment schools. If there is an odd number of schools then the last school will be assigned to treatment if a further random binary digit is 0 (rather than 1).

Incentives

Schools allocated to the intervention group will receive £700 and those allocated to the business-as-usual group will receive £1000. Incentives have been offered to both groups in order to try and reduce the potential for schools to 'dropout'. For the intervention schools, it is hoped that the incentive can facilitate the release of two English staff to attend training. For both the intervention and business-as-usual groups it is anticipated that the incentives will also encourage schools to continue in supporting the evaluation with the provision of survey data, and permitting visits and communication with participating schools. The sums will be paid in instalments following completion of certain stages/requirements of the evaluation. For both groups, 50% of the financial incentive will be paid following randomisation and a further 50% following completion of the final staff workload survey (in Spring 2019).

Participants

All state-funded mainstream secondary schools (with a Key Stage 4 cohort) in England are eligible to participate in this trial. Recruitment will be carried out by the developers, Meols Cop High School, but will be supported by the evaluation team where needed. It is expected that schools from across the country will wish to participate and the developers intend to create a number of 'hubs' in different geographical regions in order to support training and facilitation of the intervention. Schools that are interested in being part of the trial are currently being asked to contact lead developer, Sarah Cunliffe at Meols Cop High School.

Year 10 cohorts (in September 2018) in the participating schools will be the focus of this study. These same children will continue to be part of the study for two academic years i.e. until the end of their GCSE English Literature and Language courses. It is expected that teachers in schools will be using the FLASH marking approach

¹ [FLASH Marking | Projects | Education Endowment Foundation | EEF](#)

with the children over two academic years. This should allow enough time for teachers to familiarise themselves with the approach in the first year, and embed the approach in their practice in their second year. With the exception of those who opt-out of the study, all children within these cohorts will form part of the trial.

Sample size calculations

The pupil sample size calculation is based on the assumption that there will be 100 schools participating in the project. The developers have suggested having six regional 'hubs' to provide training and support for the intervention schools. If spread equally, this would mean that there would be an average of about eight schools represented in each hub. All pupils in one-year group will take part in the trial (Year 10 in September 2018) – with the exception of those who opt-out. Experience suggests an average year group of 125 pupils, meaning that there will be an approximate total of 12,500 pupils involved in the trial as a whole. Each arm of the trial will include approximately 6,250 children.

Whatever the final number of schools involved, it is vital that all cases are retained. The evaluation team are happy to attend and address recruitment events in order to explain how the trial will work and to stress the importance of committing for the duration of the evaluation even if schools do not continue with the intervention.

Traditional power calculations are based on the approach of significance testing (Gorard et al. 2017). They are not included here. Instead, we calculate the sample size needed for any 'effect' size to be considered secure by considering *a priori* the number of 'counterfactual' cases needed to disturb a finding (Gorard and Gorard 2016). This number needed to disturb (NNTD) is calculated as the 'effect' size multiplied by the number of cases in the smallest group in the comparison (i.e. the number of cases included in either the control or treatment group, whichever is smaller).

This is a useful measure of the scale of the findings to chance (and their variability as represented by the standard deviation used to compute the 'effect' size), taking into account the scale of the study. It can then be extended to compare this sensitivity directly to other more substantial sources of error such as the number of missing values/cases. The number of cases actually missing a value can be subtracted from the NNTD to give an estimate of how large the 'effect' size would be even in the extreme situation that all missing cases had the "counterfactual" score hypothesised in the NNTD calculation. Here the 'counterfactual' score is one standard deviation away from the mean of the group with the largest number of cases. The standard deviation would be added if the mean of the smaller group (in scale) were smaller than the mean of the larger group, and subtracted if the mean of the smaller group was the largest (Gorard et al. 2017).

Based on Gorard et al. 2016, NNTD of 50 can be considered a strong and secure finding. Using this as a working assumption, we would expect to detect an 'effect' size as low as 0.01 or 50/6,250 (rounded to two decimal places). The NNTD calculation concerns the security of a difference, and so is relevant to internal validity only. Issues such as clustering, concerned with whether the result may also occur among cases not in the RCT, are therefore irrelevant.

Outcome Measures

The primary outcome for this evaluation was students' attainment in GCSE English Literature and GCSE English Language. As noted above, due to the unavailability of this data, it will not be possible to respond to the primary research question regarding the academic impact of FLASH Marking.

A secondary outcome from the project focused on teacher workload, in particular workload that is related to marking and assessment. In order to measure this we will draw upon data from an online teacher workload survey completed by teachers prior to randomisation and again during the second year of the trial (prior to school closures due to Covid-19).

The survey used similar to those found in the DfE workload survey. This was most recently used in 2016 and had a total of 3,186 respondents (DfE, 2016). The survey asks teachers to self-report the amount of time that they spend on the different activities that form part of their job. The focus of the FLASH Marking survey is on issues of marking, assessment and feedback but it was also important to gain a sense of what other activities teachers spend time on in order to understand the balance of their workload, and for us to see if this alters following involvement in the FLASH marking study. A copy of the questions used in the DfE's survey can be found in the Workload Survey report (DfE, 2016). The questions used as part of the FLASH Marking survey will be published as appendices in the final evaluation report.

Fidelity measure

Fidelity to the intervention was assessed prior to schools being closed due to the Covid-19 pandemic. The following data will be used to understand fidelity and compliance but, of course, cannot be linked with any academic outcomes:

1. Number of training sessions (out of three) that staff from intervention schools attended.
2. Confirmation that cascade training was delivered to Year 10 English teachers in each school prior to trial start in September 2018.
3. Reported compliance with FLASH marking across department and for first 15 months of trial - to be asked in a question to heads of department on the teacher questionnaire in Spring 2019

Information about fidelity and compliance will be included as part of our implementation and process evaluation report.

Analysis plan

Further details of our analysis can be found in the amended statistical analysis plan (SAP).

Primary outcome

No primary outcome will be collected for this trial due to the lack of externally-assessed GCSE information from 2020 and 2021.

Secondary outcome

As noted above, we will use data from the pre/post teacher workload survey to examine the impact of FLASH Marking on English teachers' workload (as per original plan)

Attrition

For the teacher workload survey, there was substantial attrition due to teachers leaving schools or moving to new roles, or choosing not to complete the second survey. This means that we can only match teachers' responses from the pre and post surveys in some cases i.e. where they completed both surveys. Further details of our analysis are provided in the SAP.

Implementation and Process Evaluation (IPE)

The IPE was designed to provide important information about the implementation of FLASH Marking in schools and how it was received and engaged with by teachers and pupils.

At the outset of the trial a total of 15 case study schools (12 intervention, three control) were identified. Schools were selected with a view to achieving a range in terms of region, size and school intake demographics. We intended to visit the schools twice per academic year (a total of four times over the course of the two year trial). During the visits to intervention schools, the evaluation team interviewed staff involved in FLASH Marking and students about marking and assessment practices. Each visit included

interviews/discussions with at least one member of English teaching staff; where there were opportunities to speak to others, we also undertook brief interviews with these colleagues too.

Observations of teaching and feedback in Key Stage 4 English lessons were also carried out as well as examination of students' exercise books. Numbers of lessons observed per school varied depending on the timing of the visits, the teachers available and the year groups being taught at the time. In some schools we observed full lessons while in others we were invited into a number of part-lessons to see FLASH Marking or other assessment/feedback processes. The exercise book scrutiny approach relies upon us being able to access pupils' books during our visits to schools; this was possible in the majority of school visits. As pupils were often using these while we were visiting, the examination of these was very 'light touch' and predominantly involved looking to see whether FLASH Marking codes and resources were present within the books, and how these were being used by pupils and teachers. This also formed the basis for brief conversations with students about their experiences and perceptions of using FLASH Marking (or other approaches) for assessment and feedback in English.

Fieldnote data collected from the interviews and observations will be analysed using a thematic approach (Braun and Clarke, 2006). This approach offers a systematic yet flexible method of examining the data collected during the school visits. Our analyses will follow the six phases of the systematic approach (familiarisation, initial coding, identifying themes, reviewing themes, defining and naming themes, and writing up overall findings). In line with our aims of collecting rich and insightful information in order to complement and elucidate the impact data and to support our understanding of the implementation and enactment of FLASH Marking, our analysis will predominantly take an inductive approach. This data-driven approach allows us to identify themes and patterns which are not as influenced by our pre-determined research questions and aims. In addition to this though, we also include a deductive element to the analyses, using existing theory and literature to inform the identification of possible themes which may be present in the data.

The visits to control schools focused on interviewing English-teaching staff, usually through short individual interviews with available staff, about their existing marking and feedback policies and practices plus observing lessons and looking at students' workbooks. As with the intervention schools, the aim was also to speak to some young people about experiences of and attitudes towards assessment and marking. This tended to be done during English lessons which evaluation staff were invited to observe. If teachers allowed us to speak to pupils during this time, then we would use the opportunity to ask groups of pupils or individuals (depending on seating arrangements and permissions from teachers) about assessment in English. These conversations, as with the ones that took place in intervention schools, were necessarily fairly spontaneous and informal. This was to ensure that pupils felt at ease speaking to us, and did not feel the need to 'prepare' what they wanted to say or to try and present a particular view on the issues raised. We were also very conscious that students were participating in important lessons, preparing them for two GCSEs in a core subject (English). As such, all interactions were kept relatively brief so that precious learning time was not missed.

The process evaluation also includes a brief staff attitudes survey to explore English teachers' views on marking and feedback. Questions relating to these topics were included within the second teacher workload survey in order to reduce the need to ask staff to complete a separate questionnaire. These questions were available to both intervention and control group teachers. The attitudes questions were drawn up in conjunction with the development team. For analysis, means and standard deviations will be provided for 'real number' responses such as hours worked per kind of activity, and percentages for all categories such as methods of assessment and views on workload and FLASH Marking. We will compare the means between treatment and control responses and convert to standardised mean differences. Frequencies between treatment and control responses will also be cross-tabulated.

A short student survey was also carried out with a small number of intervention schools who were willing to participate. This was administered to schools during the second year of the trial and could either be completed as a paper copy or online. The aim of the survey was to examine issues of awareness, responsiveness, reach, and programme differentiation. Through the survey we also hoped to gain an understanding of children's attitudes and experiences in relation to FLASH Marking and marking/feedback more generally. The student survey was targeted at FLASH intervention schools. The original aim was to involve all of the case study schools. A total of eight of the 12 intervention case study school participated. Two schools declined to participate due to the time that would be required for students to complete the survey. Two schools agreed to participate in the Spring term of 2020 but, with the arrival of the pandemic and school closures, felt unable to do so. All survey instruments will be included as appendices in the final report.

In overview, the data collected as part of the implementation and process evaluation aimed to assess:

- teachers' response to training, including attendance rates
- the fidelity of training
- teachers' delivery of the intervention
- whether the teams understand the process and purpose
- the contents and use of any materials
- changes in teacher behaviour
- staff and students' views of the intervention
- any apparent impact on children's behaviour and attitudes
- possible indication of contamination or diffusion

Further details about the analysis of the IPE data can be found in the amended SAP.

Costs:

As proposed in the original protocol, the costs to schools for implementing the intervention will be determined. These will be calculated per pupil using the following estimates:

Cost of setting-up

- Cost of delivering training to English teaching staff
- Cover for teachers to attend the three training sessions
- Any other costs

Cost of delivery

- Teaching materials and resources associated with FLASH marking
- Day rates for teaching staff
- On-going monitoring and support from the delivery team

Other non-monetary costs

Time taken away from regular lessons for attending training, organising the administration and collection of the training/cascading materials. Once set up, it is not anticipated that staff will lose time in the classroom. Indeed, it is possible that marking time will be reduced for staff and, if this is the case, we are interested in how this time is being used instead. Information on this will be collected with input from the project team and via surveys and interviews with teaching staff.

Ethics

A number of ethical considerations will be taken into account during the planning, implementation and reporting stages of the trial. The BERA ethical guidelines (2018) and Durham University's strict research Code of Practice will be adhered to at all times. Ethical approval has been gained from Durham University's Research Ethics Committee.

Parental consent will be sought in relation to the use of students' data in the analysis and reporting of the trial and outcomes. This will involve an 'opt out' consent form and information letter whereby parents and carers can choose to have their child's information withdrawn from the final analyses and reporting.

All schools, staff and pupils involved in the trial will remain anonymous. Participant information will be treated confidentially and all participants in interviews and observations will be informed of their right to withdraw at any stage. No person or school will be identifiable in the reporting of this trial.

The GCSEs exams will be occurring irrespective of this trial and so will add no additional burden in terms of testing.

The data used from the National Pupil Database will contain potentially sensitive information about the children involved in the trial. For the purposes of this project, however, we will only need Tier 2 data (rather than the more sensitive Tier 1 information), in order to conduct the analyses detailed above. No pupils will be identifiable and all secondary data will be kept in secure environments, following the strict data protection requirements of the NPD and Durham University.

Durham University's data protection policy is publicly available at: <https://www.dur.ac.uk/ig/policies/dppolicy/>. The privacy notice for this project can also be found here: <https://www.dur.ac.uk/research/directory/view/?mode=project&id=1014>

The Ethics Committee at Durham University have been informed of the recent changes to the evaluation and the removal of the primary outcome and data analysis.

Personnel

Evaluation Team – School of Education, Durham University

Dr Rebecca Morris (at University of Warwick from January 2019) will be responsible for the final delivery of all outputs and meeting deadlines. She will lead on the day-to-day organisation of the study, arranging fieldwork, communicating with the EEF, collecting data for the impact and process evaluations, analysis and report writing.

Prof. Stephen Gorard will be responsible for the design and analyses of the impact evaluation, and will assist with all other elements, including report writing.

Dr Beng Huat See will support with all aspects of the fieldwork, data collection, analysis and report writing. She has led a number of previous EEF evaluations, and her work is dedicated to such projects and other similar opportunities.

Dr Nadia Siddiqui will assist with fieldwork, data collection and cleaning, arranging fieldwork, and assist with communicating with the study researchers, analysis and report writing.

Research assistants will be employed for parallel fieldwork and to relieve pressure on principal researchers. Duties will include cleaning and preparing data, data coding, fieldwork and literature searches.

Project team – Meols Cop High School, Southport

The intervention team from Meols Cop High School will be responsible for school recruitment, collecting opt-out consent from parents, ongoing relationships with schools and keeping parents informed of the intervention (if necessary). They will be responsible for the training of staff, supporting the cascading of training and the delivery of the intervention. They will work with schools to collect baseline data. These tasks

will be conducted with the support of the evaluation team.

Timeline: the chart below outlines the activities as part of the evaluation. Those highlighted in red refer to activities that were cancelled or delayed due to the Covid-19 pandemic. On the following page, we provide details of the timetable for completing the revised evaluation.

	2017-2018				2018-2019				2019-2020						2020-2021					
	Spring term		Summer term		Autumn term		Spring term		Summer term		Autumn term		Spring term		Summer term		Autumn term		Spring term	
Recruit schools																				
Collect pre-test data from schools																				
English teachers complete first workload survey																				
Randomise schools																				
Confirm teachers and class lists																				
First training session and cascading																				
Teachers use FLASH marking with Year 10 classes																				
Second training session and cascading																				
Classroom observations and interviews																				
Teachers continue using FLASH marking																				
Third training session																				
Second teacher workload survey																				
Classroom observations and interviews																				
Year 11 students take GCSE Exams																				
Collect post-test data																				
Analysis and report writing																				
Final report complete																				

Timetable for completion of FLASH Marking evaluation (2020-2022)

	2020-2021				2021-2022	
	Spring term		Summer term		Autumn Term	Spring term
DfE decision to cancel GCSEs for 2021.						
Decision to revise evaluation and remove primary outcome.						
Collation and analysis of data (teacher survey, IPE).						
Report writing						
Submission of first draft to EEF (September 2021)						
Submission of final draft to EEF (February 2022)						

References

- BERA (2018) Ethical Guidelines for Educational Research. Available: <https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018>
- Black, P., and Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. London: Granada Learning.
- Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), pp.77-101.
- Christodoulou, D. (2017) Making good progress? The future of assessment for learning, Oxford: Oxford University Press
- DfE (2016) Teacher Workload Survey 2016, DfE: London. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/592499/TWS_2016_FINAL_Research_report_Feb_2017.pdf
- Elliot, V., Baird, J., Hopfenbeck, T., Ingram, J., Thompson, I., Usher, N., Zantout, M., Richardson, J., Coleman, R. (2016) A marked improvement? A review of the evidence on written marking, London: EEF
- Gibb, N. (2015) Assessment after levels. Speech, 15th February 2015. Available from: <https://www.gov.uk/government/speeches/assessment-after-levels>
- Gibson, S., Oliver, L. and Dennison, M. (2015) Workload Challenge: analysis of teacher consultation responses research report. London: Department for Education. Available from: www.gov.uk/government/uploads/system/uploads/attachment_data/file/401406/RR445_-_Workload_Challenge_-_Analysis_of_teacher_consultation_responses_FINAL.pdf
- Gov.uk (2016) Ofsted Inspection Myths <https://www.gov.uk/government/publications/school-inspection-handbook-from-september-2015/ofsted-inspections-mythbusting>
- Hattie, J. and Timperley, H. (2007) The Power of Feedback, *Review of Educational Research*, 77(1): 81-112
- Klapp, A. (2015). Does grading affect educational attainment? A longitudinal study. *Assessment in Education: Principles, Policy and Practice*, 22(3), pp. 302-323
- Lilly, J., Peacock, A., Shoveller, S. and Struthers, D. (2014) Beyond Levels: alternative assessment approaches developed by teaching schools Research Report, London: DfE. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/349266/beyond-levels-alternative-assessment-approaches-developed-by-teaching-schools.pdf
- Lipnevich, A. and Smith, J. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, 15(4), pp. 319-333
- Ofqual (2015) GCSE changes: a summary. Available from: <https://www.gov.uk/government/publications/gcse-changes-a-summary>
- Smith, E. and Gorard, S. (2005) 'They don't give us our marks': the role of formative feedback in student progress, *Assessment in Education*, 12 (1), pp.21-28
- Zhang, B. and Misiak, J. (2015). Evaluating three grading methods in middle school science classrooms, *Journal of Baltic Science Education*, 14(2), pp.207-21

