

Supporting teachers through the school workload reduction toolkit

Richard Churches

Education Development Trust

March 2020

Contents

List of tables and figures	3
Executive summary	4
Project outline and approach	4
Findings	4
1. Methodology	6
1.1 How the project was conducted: Training and support	6
1.2 The measures used by teachers	8
2. Teacher findings	10
2.1 Teacher time	10
2.2 Teacher wellbeing	11
2.3 Pupil attainment and progress	12
2.4 Differences in impact on pupil attainment and progress	14
3. Conclusions	16
Appendix 1 – International Personality Item Pool scale results	22
Appendix 2 – Forest plots of pupil attainment and progress	23
Forest plots and how to read them	23
Teacher research designs included in the meta-analysis	30

List of tables and figures

Figure 1: Teachers sharing findings at the celebration event	7
Figure 2: A teacher-led workload reduction conference poster	8
Figure 3: Teacher average time on target tasks (with and without workload reduction)	10
Figure 4: International Personality Item Pool scores	12
Table 1: Separate Mann-Whitney U test IPIP results (control versus intervention)	22
Figure 5: Communication workload reduction interventions (forest plot)	24
Figure 6: Data recording and reporting workload reduction interventions (forest plot)	25
Figure 7: Lesson observation and monitoring workload reduction interventions (forest plot)	26
Figure 8: Marking and feedback workload reduction interventions (forest plot) Part 1	27
Figure 9: Marking and feedback workload reduction interventions (forest plot) Part 2	28
Figure 10: Marking and feedback workload reduction interventions (forest plot) Part 3	29

Executive summary

Project outline and approach

Reducing teacher workload can make an important contribution to improving teacher retention and wellbeing. The government has been supporting school-led strategies that address the issue. However, the potential wider effects of reducing teacher activity in areas such as marking, planning and data recording on pupil attainment and progress are less understood.

The Department for Education (DfE) provided funding for the Teaching Schools Council (TSC) to promote the School Workload Reduction Toolkit and support schools to use it in ways that worked in their own contexts. The TSC representatives encouraged schools from across England to use the toolkit to identify priority areas and adopt new practices to address them. Education Development Trust was commissioned to support schools in assessing the impact of any changes they made. Teachers designed studies to look at the effects of a wide range of workload reduction strategies on three key elements:

- teacher time, spent on activities such as cross-school communications, collating and reporting pupil data, lesson planning and monitoring, and feedback and marking
- teacher wellbeing and
- pupil learning outcomes.

Most studies took place over one term and investigated areas such as marking and feedback, lesson planning, monitoring, data reporting and communication policies. Despite the short length of time over which the studies were conducted, the approach generated useful findings that illustrate how workload can be significantly reduced without having a detrimental effect on pupil outcomes, while at the same time improving teacher wellbeing.

Findings

Teacher designed interventions significantly reduced teacher time conducting the targeted tasks, i.e. approaches to marking and feedback, lesson planning, managing pupil data, internal communications, and lesson observation and monitoring.

Where schools measured wellbeing, using valid and reliable scales from the International Personality Item Pool (IPIP)¹, **wellbeing overall improved**. There were

¹ Goldberg et al. (2006)

significant reductions on the workaholism scale² (the IPIP scale used to assess the extent to which teachers were working too hard) and significant increases in self-efficacy³ (a personal judgment of "how well one can execute courses of action required to deal with prospective situations"⁴).

Across the studies overall, **reducing teacher workload was associated with a period of maintained or improved pupil outcomes**. For alternative marking and feedback approaches that provided individual feedback in the classroom, there may have been attainment and progress improvements alongside the reductions in teacher workload outside of class. Effective strategies included immediate formative assessment and teaching pupils metacognitive strategies to plan, monitor and evaluate how well they learn.

² Simms et al. (2011)

³ Costa and McCrae (1992)

⁴ Bandura (1982)

1. Methodology

The school workload reduction toolkit is a set of practical resources for school leaders and teachers that aims to help reduce workload, published in July 2018. It was produced by school leaders, teachers and other sector experts together with the Department for Education⁵ drawing on successful workload reduction strategies undertaken by schools. This included conducting case study research to further explore and develop a range of approaches. The toolkit was updated in October 2019 with more sections and a revised navigation. In 2019 the Department for Education commissioned Education Development Trust to supply training in school-based research methods across 8 Teaching School Council regions, to better understand the impact of the workload reduction toolkit in schools:

- East of England and North East London
- East Midlands, South Yorkshire and The Humber
- Lancashire and West Yorkshire
- North
- North West London and Central England
- South East England and South London
- South West
- West Midlands

The training consisted of three face-to-face days with remote support between the training days and at the end of the project⁶. There was a celebration and evaluation event in London in the middle of March 2020 allowing teachers to meet each other and share the outcomes of their studies (Figure 1).

1.1 How the project was conducted: Training and support

The training and support given to schools consisted of a mixture of face-to-face training events and remote support:

1. **Training day 1** (May/June 2019). This day introduced teachers to a variety of research methodologies they could use in their studies if they wanted to. They were taught how to design randomised controlled trials, non-randomised controlled trials and retrospective cohort studies - involving either a control comparison group of

⁵ <https://www.gov.uk/government/collections/reducing-school-workload>

⁶ This approach has been effective on several previous occasions, as part of a Department for Education programme (Closing the Gap: Test and Learn) and in teacher-led research funded by the Wellcome Trust and the Varkey Foundation (Churches, 2016; Churches and Dommett, 2016; Churches et al., 2020; Churches, Higgins and Hall, 2018; Churches, Korin and Sims, 2020).

pupils, or in the form of a uncontrolled retrospective analysis comparing the progress/attainment of a group of pupils in this current academic year compared to the same time period last year.

To help develop their research design and focus, teachers used the staff workload survey questionnaire⁷ from the school workload reduction toolkit and looked at previously completed teacher workload case studies.

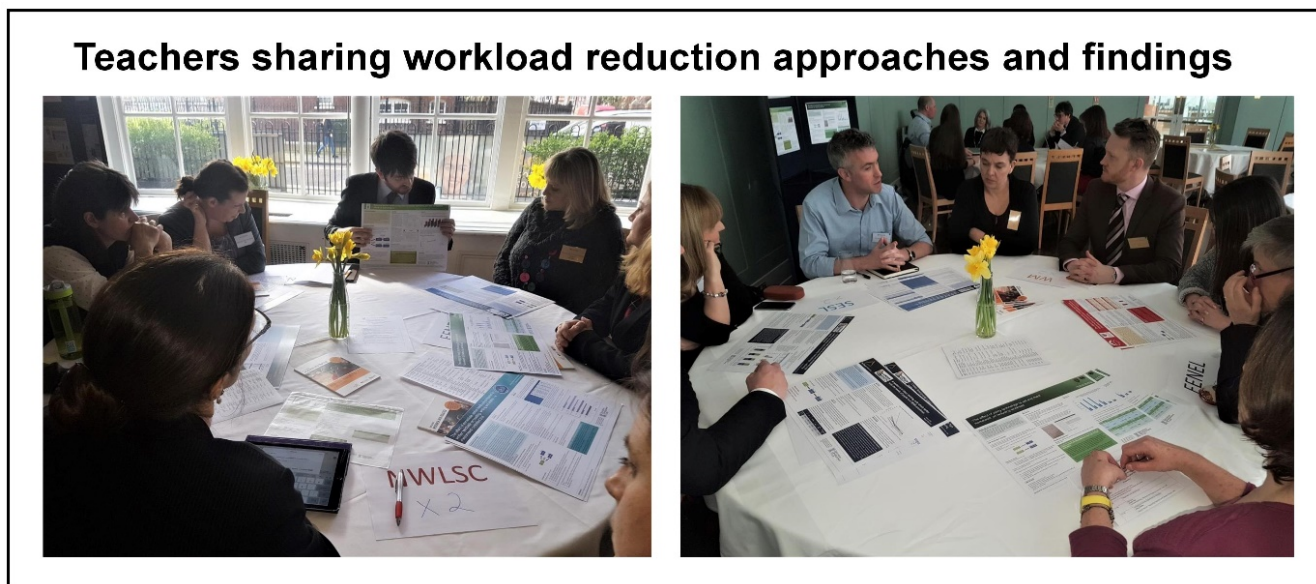


Figure 1: Teachers sharing findings at the celebration event

Source: Education Development Trust

2. **Research protocol sign off and feedback** (July 2019). Following the first training day, teachers completed a research design protocol template and sent these in for remote feedback.

Of the 42 research protocols received from 7 of the 8 regions, teachers implemented 36. These 36 included 112 separate trials in different year groups or subjects, measuring pupil attainment or progress.

3. **Training day 2** (November/December 2019). This training day taught teachers how to analyse their results and draft a research conference style poster of their results which they could be share with other teachers⁸. The training included understanding and interpreting the results and how to conduct statistical analyses.

⁷ <https://www.gov.uk/guidance/school-workload-reduction-toolkit>

⁸ Teaching Schools received a research methods textbook to support their work (Churches and Dommett, 2016). Analyses used Education Development Trust's EXCEL analysis programmes. StatsWizard

- 4. Training day 3 (January to March 2019).** Day 3 gave teachers time to finish their conference posters (see Figure 2). They completed their analyses and began to draft summary text to support the dissemination of their findings.
- 5. Conference poster remote support and feedback (February/March 2020).** On completion of their analyses and writing up, Education Development Trust gave teachers support and feedback to help them complete their conference posters (Figure 2).

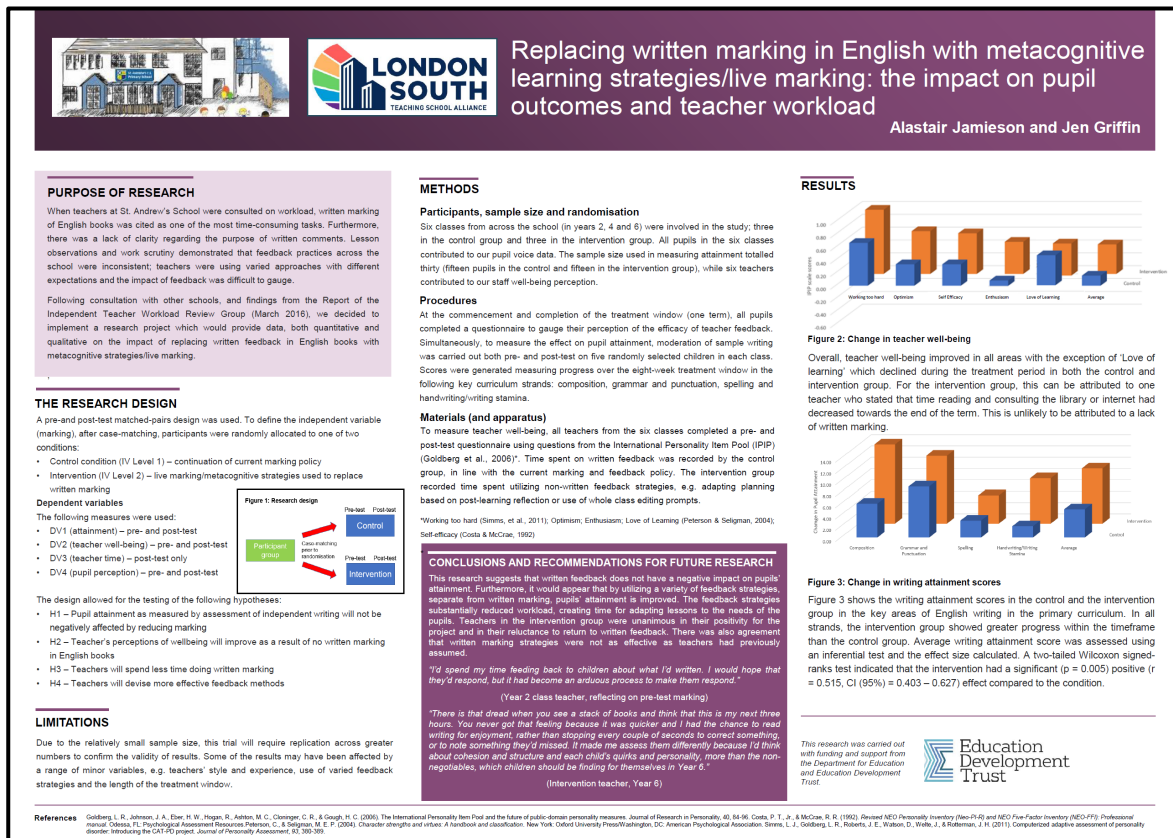


Figure 2: A teacher-led workload reduction conference poster

Source: Jamieson and Griffin (2020)

1.2 The measures used by teachers

The project asked schools to measure the following areas (where possible):

- **Teacher time** – average minutes per week conducting the tasks targeted for workload reduction.

automatically conducts assumption testing, directs teachers to the correct inferential test, calculates effect sizes, confidence intervals and p-values and drafts a preliminary results paragraph.

- **Teacher wellbeing and perception** - through teacher designed questionnaires. In addition, teachers received an optional questionnaire constructed from pre-selected International Personality Item Pool (IPIP)⁹ scales.
- **Attainment/progress** – using their normal assessment data and usually with a ‘does no harm hypothesis’ (i.e. there will be no change in pupil outcomes).

Some teachers also collected pupil perception data and conducted qualitative interviews/focus groups.

⁹ The International Personality Item Pool (IPIP) (Goldberg et al., 2006) is a website (<https://ipip.ori.org/>) that includes over 3,000 items and over 250 scales. All the items and scales are in the public domain, meaning that researchers can use them for any purpose without permission or paying a fee.

2. Teacher findings

Amalgamating teacher individual study findings, we were able to analyse national data in the following three areas.

2.1 Teacher time

Where schools had measured teacher time they sent in their anonymised data. Teacher time data was available for 267 teachers from 14 of the teacher-led studies. **Workload reduction interventions reduced teacher average time on the target tasks from around 1 hour and 20 minutes to half an hour** (Figure 3). In addition, variation in teacher working time reduced in the interventions,¹⁰ suggesting that the teachers now had more similar levels of workload. There was a moderate significant effect (equivalent to $d = -0.59$)¹¹.

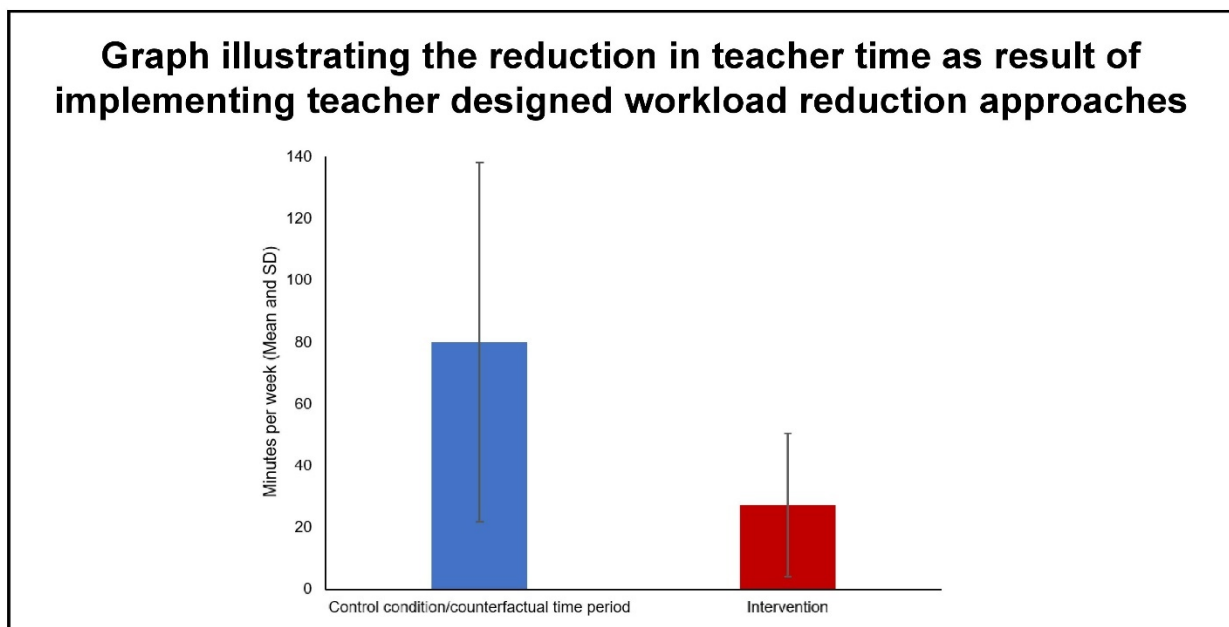


Figure 3: Teacher average time on target tasks (with and without workload reduction)

Source: Data from 14 teacher-led research projects (N = 267)

¹⁰ M = 80.02, SD = 109.10 (control condition/counterfactual period); M = 27.36, SD = 38.94 (intervention)

¹¹ A one-tailed Mann-Whitney test indicated that teacher time on task was significantly reduced by the workload reduction approaches (Mdn = 11) compared to the control conditions and comparison periods (Mdn = 50), $Z = 8.59$, $r_{pb} = -0.26$, $p < .001$, CI (95%) = -0.439 – -0.132.

Teacher designed interventions were effective in reducing and equalising teacher time on the targeted tasks

- There was a significant reduction in teacher time across the studies that measured this area.
- The difference between the teachers spending the most time and those spending the least time working on relevant activities also reduced.

2.2 Teacher wellbeing

Six schools measured staff wellbeing, using the 5 International Personality Item Pool (IPIP) scales shared during the research design training days:

- Workaholism (used to measure the extent to which teachers were working too hard)
- Optimism
- Self-efficacy (belief in one's ability)
- Enthusiasm
- Love of learning¹²

Workload reduction approaches chosen by the teachers had an overall significant effect on teachers' wellbeing across all 5 measures¹³. There were significant positive effects on working too hard (as measured by the workaholism scale), self-efficacy¹⁴ and on the combined average of all scales¹⁵.

¹² All scales had acceptable levels of internal consistency in prior research: Workaholism ($\alpha = .83$; $\alpha = .85$); Optimism ($\alpha = .80$); Self-efficacy ($\alpha = .78$); Enthusiasm ($\alpha = .78$); Love of learning ($\alpha = .77$) (see <https://ipip.ori.org/>). Internal consistency for the present analysis data was also acceptable (see Appendix 1, Table 1).

¹³ An initial Kruskal-Wallis ANOVA across the change in scores for the 5 International Personality Item Pool scales indicated a significant large effect size difference across all of the variables, $H(4) = 9.51$, $p = .049$, $\eta_p^2 = 0.137$ [$d = 0.80$], suggesting that the effects were unlikely to be the result of family-wise error. We then conducted separate Bonferroni-adjusted Mann-Whitney U tests on each of the variables and the combined average scores (see Appendix 2, Table 1).

¹⁴ $\alpha = 0.01$ (Bonferroni-adjusted)

¹⁵ $\alpha = 0.05$

Teacher wellbeing improved with teacher designed interventions

- Overall teacher wellbeing improved in the teacher-led research that measured this area.
- There were significant improvements in whether the teachers were working too hard and feelings of self-efficacy.

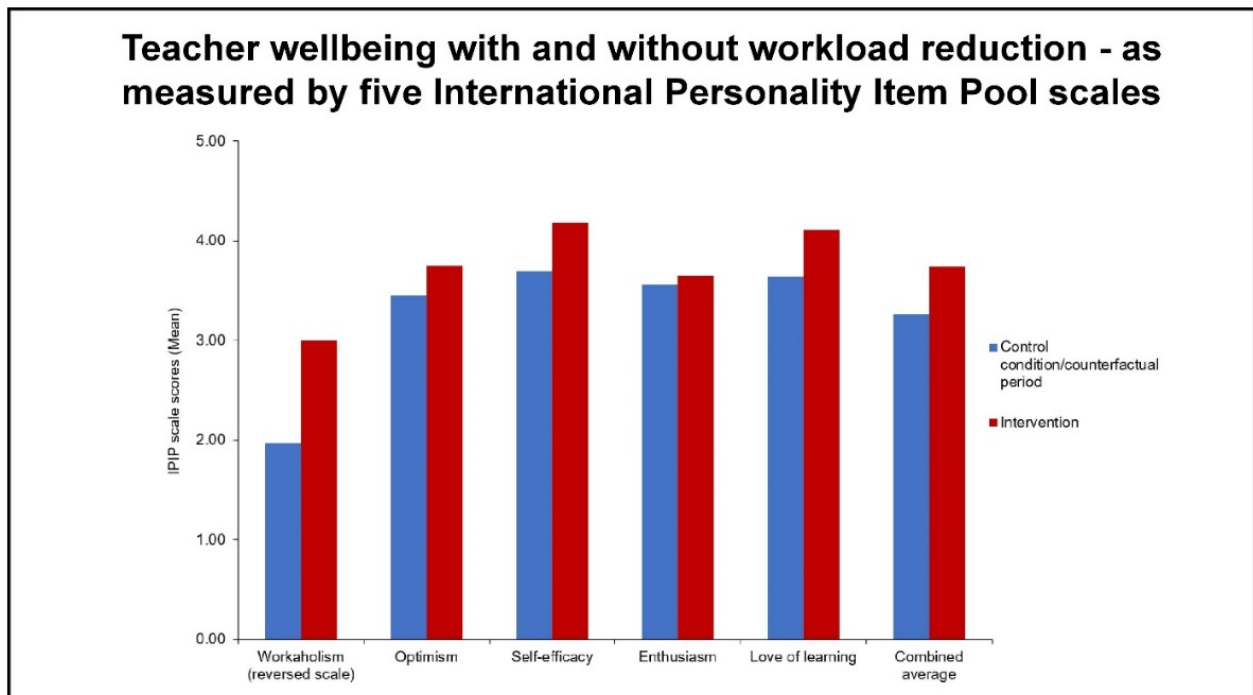


Figure 4: International Personality Item Pool scores

Source: Data from 6 teacher-led research projects (N = 24)

2.3 Pupil attainment and progress

All schools measured attainment or progress¹⁶ as they were able to use existing teacher/school data in most cases. As many of the schools taking part replicated the approaches with different year groups and in different subjects, we were able to

¹⁶ Teacher research protocols included a mixture of post-test only and pre- and post-test data. Where studies had pre- and post-test interval data the teachers calculated gain scores and compared progress rates using these scores. Many schools, particularly primary schools, have moved to data which categorises pupils into levels of on-track or off-track progress. In these cases, teachers used chi-squared tests of independence, with the effect sizes calculated from the test statistic. This type of data presents challenges within a meta-analysis because on occasions the effect size can be exactly 0.00 with a CI of exactly 0.00 – 0.00. In these cases, we estimated the CI from other available data.

calculate 112 effect sizes¹⁷ measuring the strength and direction of changes in pupil attainment/progress from 25 teacher research protocols and 10,980 pupil level assessments. The Education Development Trust conducted a meta-analysis¹⁸ of these effect sizes to find the average effect of introducing new working practices based on ideas in the School Workload Reduction Toolkit. Effects were organised into 4 areas to enable sub-group meta-analyses:

- Communication
- Data recording and reporting
- Lesson observation and monitoring
- Marking and feedback

Across all the teacher study effect sizes¹⁹ workload reduction was associated with a small but significant positive effect²⁰. 90.70% of effect sizes were positive suggesting that pupil attainment and progress remained consistent or improved during the workload reduction period, compared to the control conditions and retrospective comparison periods. Over one-third of findings were significant (39.20%).

Sub-group meta-analyses indicated that an overall moderate significant positive effect was associated with the period of reduced workload for the communication interventions²¹, while small significant positive effects were associated with data recording²² and marking and feedback²³. There was a small non-significant positive effect associated with the period of workload reduction in areas related to lesson observation and monitoring²⁴. Appendix 2 contains the forest plots for the sub-group meta-analyses (Figures 6 to 10).²⁵

¹⁷ Effect sizes describe the strength and direction of any difference between the intervention and control conditions. The commonest effect size reported in education is Cohen's d (see for example, Hattie, 2009; 2012). A Cohen's d of 0.5 is considered a moderate positive effect (and equates to an approximate 33% non-overlap between the intervention and control scores), a Cohen's d of -0.2 would be considered a small negative effect (an approximate 14.7% non-overlap).

¹⁸ Random effects (using Suurmond, van Ree and Hak, 2017). See Higgins (2018) for a discussion of meta-analysis in education.

¹⁹ Because many of the trials and quantitative analyses produced data that was non-parametric, all effect sizes have been reported using r (Rosenthal, 1991). Where we have done this, we have also given the Cohen's d equivalent.

²⁰ (Overall) $r = 0.11$ [$d = 0.22$], CI (95%) = 0.06 – 0.16), $Q = 1527.13$, $p < .0005$, $I^2 = 92.73\%$, $T^2 = 0.02$, $T = 0.15$

²¹ (Communication) $r = 0.19$ [$d = 0.39$], CI (95%) = 0.12 – 0.26, $Q = 76.81$, $p < .0005$, $I^2 = 73.96\%$, $T^2 = 0.02$, $T = 0.14$

²² (Data reporting) $r = 0.11$ [$d = 0.22$], CI (95%) = 0.003 – 0.21, $Q = 1047.17$, $p < .0005$, $I^2 = 98.66\%$, $T^2 = 0.04$, $T = 0.20$

²³ (Marking/feedback) $r = 0.11$ [$d = 0.22$], CI (95%) = 0.08 – 0.14, $Q = 363.90$, $p < .0005$, $I^2 = 80.46\%$, $T^2 = 0.01$, $T = 0.11$

²⁴ (Lesson observation/monitoring) $r = 0.06$ [$d = 0.12$], CI (95%) = 0.02 – 0.11, $Q = 2.72$, $p = .437$, $I^2 < 0.001\%$, $T^2 < 0.001\%$, $T < 0.001\%$

²⁵ Future similar programmes could use of meta-regression to moderate contextual factors with a larger number of replications, perhaps involving 15 or more studies coded for the same moderating variables (see Valentine, Pigott and Rothstein (2010) for a discussion of the power-related issues that can arise during the meta-analysis of small-scale studies).

Meta-analysis of pupil outcome data suggests reducing teacher workload does not affect pupil progress or attainment and may improve outcomes.

Overall pupil attainment and progress remained constant, or improved, during the intervention of teacher reduced workload. In some cases (such as certain alternative marking and feedback strategies), reducing workload outside of class may in fact improve attainment.

2.4 Differences in impact on pupil attainment and progress

It is clear from the evidence in the forest plots that there were few negative effects on pupil attainment and progress during the project. Although, we cannot assume the effects are entirely the result of the strategies that schools chose. Where there were negative effects, schools were mostly able to explain the causes of these in relation to other contextual or pupil factors.

This is an important finding for the teacher profession and for policy implementation. The need to reduce workload has become clear because of the negative effects on teacher retention and wellbeing from high workload. However, headteachers and Trust leads may find themselves cautious to remove or reduce processes which they believed to be important for maintaining or improving the school's academic performance - such as detailed lesson planning, extensive written marking outside of class and frequent data reporting.

The pattern of results across the areas studied by the teachers can be summarised as follows:

- **Communication.** Two teacher whole school research designs generated 21 effects. The research designs mirrored each other and explored more efficient ways to communicate daily messages and other forms of communication across the schools. All but 2 effect sizes were positive (Appendix 2, Figure 5).
- **Data reporting.** Two research designs produced 15 effect sizes that looked at attainment and progress during periods where the number of 'data drops' was reduced across the schools (Appendix 2, Figure 6). Similarly, the majority of pupil outcomes were positive.
- **Lesson planning and monitoring.** Two studies (1 a within-participant randomised controlled trial) produced 4 effect sizes, all were positive (Appendix 2, Figure 7).
- **Marking and feedback.** The largest number of studies and effect sizes were generated by teacher research that was interested in alternative approaches to traditional out of class written marking (with 72 effect sizes able to be included in the meta-analysis). As well as having the largest number of effects, the variation

in effect was the greatest in this area. Although again, most effects were positive. This area of exploration also generated the largest positive effect sizes as well as notable negative effects.

The strategies that involved direct individual feedback in the classroom as the children were learning were most likely to be associated with the largest effects (Appendix 2, Figures 8 to 10). This is perhaps not surprising as the real time process of direct feedback, correction of misconceptions, setting of targets and selection of strategies in response to such feedback is likely to trigger processes associated with pupil metacognition (planning – monitoring – evaluation)²⁶, an area associated with moderate to large positive effects across the large-scale randomised controlled trial literature²⁷.

In contrast, at the negative effect end of the forest plot whole class feedback approaches (as a replacement for written marking) appeared to be less effective, although there were questions within these studies as to whether teacher assessment data was reliable. However, there was variation in effect for all approaches. Further research will be necessary to unpack which form of feedback might be most effective with which children in which subject and school context.

²⁶ Churches, Domett and Devonshire (2017)

²⁷ Elliot Major and Higgins (2019); <https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/meta-cognition-and-self-regulation/>

3. Conclusions

In relation to the teacher data that we could amalgamate and analyse at a national level, over one academic term:

- There was a significant reduction in teacher time. In addition, the difference in working time between teachers who conducted the workload reduction strategies reduced.
- Teacher wellbeing improved. There were significant improvements in respect of whether the teachers were working too hard and feelings of self-efficacy.
- Overall pupil attainment and progress remained constant or improved. In some cases (such as certain alternative marking and feedback strategies), reducing workload outside of class may in fact improve attainment.

References

Ainsworth, A. (2020) Code marking in combination with flash marking reduces teacher workload and may have a positive effect on pupil writing progress. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Amos, N. (2020) The effect of in lesson feedback compared to written marking – a randomised controlled trial. Conference poster (unpublished).

Bandura, A. (1982) Self-efficacy mechanism in human agency. *American Psychologist*. 37 (2): 122–147.

Boroughs, L. (2020) The effect of implementing simplified and reduced internal school communication on teachers' perception of workload and wellbeing. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Churches, R. (2016) *Closing the Gap: Test and Learn*. London: Department for Education/National College for Teaching and Leadership.

Churches, R. and Dommett, E. (2016) *Teacher-led research: designing and implementing randomised controlled trials and other forms of experimental research*. Camarthen: Crown House Publishing.

Churches, R., Dommett, E. and Devonshire, I. (2017) [Foreword by Susan Greenfield]. *Neuroscience for Teachers: applying research evidence from brain science*, Camarthen: Crown House Publishing.

Churches, R., Dommett, E., Devonshire, I, Hall , R., Higgins, S. and Korin, A. (2020) Translating laboratory evidence into classroom practice with teacher-led randomised controlled trials - a perspective and meta-analysis. *Mind, Brain and Education*. <https://doi.org/10.1111/mbe.12243>

Churches, R., Higgins, S. and Hall, R. (2018) The potential of teacher-led randomised controlled trials in education research. In A. Childs and I. Menter (eds.), *Mobilising teacher researchers: challenging educational inequality* (pp.113–119). Abingdon: Routledge.

Churches, R., Korin, A. and Sims, K. (2020) *Test and learn: a global revolution in teacher-led research*. Reading: Education Development Trust.

Clark, S. and Cullen, K. (2020) The effect of immediate verbal feedback compared to written feedback on reading, writing and maths – a randomised controlled trial.

Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Cole, A. and Mahhire, R. (2020) A system of in class marking only, which better suits the needs of pupils in a SEND provision. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Costa, P.T. and McCrae, R.R. (1992) *Revised NEO Personality Inventory (Neo-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual*. Florida: Psychological Assessment Resources.

Davis, B. and Woodley, D. (2020) Reducing teacher workload may improve teacher wellbeing and has no negative effects on pupil progress. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Elliot Major, L. and Higgins, S. (2019) *What works? Research and evidence for successful teaching*. London: Bloomsbury.

Figueiredo, J. (2020) An uncontrolled cohort study looking at the effect of written feedback and marking on staff wellbeing and pupil outcomes. Conference poster (unpublished).

Frounks, K. (2020) Does reducing written feedback to pupils have a negative impact on progress whilst having a positive impact on teacher time and wellbeing? Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Geiger, T. and Pivovarova, M. (2018) The effects of working conditions on teacher retention. *Teachers and Teaching*, 24(6), 604–625.

Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R. and Gough, H.C. (2006) The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 849–856.

Hammond, T. and Watson, R. (2020) The effect of using technology to set and mark homework on reducing workload. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Hattie, J. (2009) *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Abingdon: Routledge.

Hattie, J. (2012) *Visible learning for teachers*. Abingdon: Routledge.

Hedley, P. and Wardle, C. (2020) The feasibility of using whole-class feedback to address common pupil misconceptions – a randomised controlled trial (geography).

Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Higgins, S. (2018) *Improving Learning: Meta-analysis of intervention research in education*. Cambridge: Cambridge University Press.

Hooper, M. and Wines, A. (2020) Reducing teacher workload improves teacher wellbeing (particularly 'Love of Learning') and has no negative effects on pupil attainment. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Jamieson, A. and Griffin, K. (2020) Replacing written marking in English with metacognitive learning strategies/live marking: the impact on pupil outcomes and teacher workload. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Knipe, S. (2020) 'No More Marking' (comparative judgement) improves pupil progress and reduces teacher workload, results from a non-randomised controlled trial. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Leech, A. (2020) The effect of implementing simplified and reduced internal school communication on teachers' perception of workload and wellbeing. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Manhire, R. and Cole, A. (2020) A holistic developmental system of ensuring a consistently high quality of education using a coaching methodology – reducing teacher workload. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

McCormack, K. (2020) Flash and self-marking reduces workload for teachers, pupils feel more successful – a preliminary study. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

McKeever, C. and Collins, C. (2020) The effect of live feedback compared to out of class marking. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Pain, M. and Soanes, E. (2020) A randomised control study looking at the impact of live marking in KS4 science. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Peterson, C. and Seligman, M.E.P. (2004) *Character strengths and virtues: A handbook and classification*. New York: Oxford University Press.

Pyne, J. (2020) Marking at the point of learning – preliminary reporting from a non-randomised controlled stepped wedge design Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Reakes, G. (2020) Assessing the effect of reducing frequency of ‘data drops’: a retrospective quantitative analysis. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Rigby-Beckett, B. (2020) Mark less, mark better – a non-randomised control trial. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Riley, R.D, Higgins, J.P.T. and Deeks, J.J. (2011) Interpretation of random effects meta-analyses. *British Medical Journal*, 342, d549.

Rosenthal, R. (1991) *Meta-analytic procedures for social research*. California: Sage

Shute, M. (2020) The effects of reducing the frequency and intensity of data drops on pupil progress, teacher wellbeing and teacher time. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Simms, L.J., Goldberg, L.R., Roberts, J.E., Watson, D., Welte, J. and Rotterman, J.H. (2011) Computerized adaptive assessment of personality disorder: Introducing the CAT-PD project. *Journal of Personality Assessment*, 93, 380–389.

Sunter, J. (2020) The effect of replacing marking with digital mastery. Conference poster (unpublished).

Suurmond, R., van Rhee, H. and Hak T. (2017) Introduction, comparison and validation of Meta-Essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods*, 8(4), 537–553.

Trevelan, T. and Minnott, K. (2020) Class big books – reducing marking in science in Key Stage 2. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London 12 March 2020.

Valentine, J.C., Pigott, T.D. and Rothstein, H.R. (2010) How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, 35(2), 215-247.

Wardle, C. and Hedley, P (2020) The feasibility of using whole-class feedback to address common pupil misconceptions (physics) – a retrospective analysis. Conference poster presented at the Department for Education Reducing Workload Celebration Event, London, 12 March 2020.

Appendix 1 – International Personality Item Pool scale results

	U	p-value	Effect Size r^{28}	[d]	Cronbach's alpha
Workaholism	124.0	.002**	0.62	1.57	.80
Optimism	94.0	.100	0.26	0.54	.76
Self-efficacy	119.0	.003**	0.55	1.34	.81
Enthusiasm	75.50	.420	0.04	0.08	.78
Love of learning	110.0	.14	0.45	1.00	.82
Combined	112.0	.01*	0.47	1.07	

**significant with a Bonferroni-adjusted alpha = .01; *significant with alpha = .05

Table 1: Separate Mann-Whitney U test IPIP results (control versus intervention)

Source: Data from 6 teacher-led research projects (N = 24)

²⁸ $r = Z/\sqrt{N}$

Appendix 2 – Forest plots of pupil attainment and progress

Forest plots and how to read them

Each dot [•] represents the effect size (Figures 5-10). This illustrates the strength and direction of any change. Error bars [|] (either side of the effect size) show 95% confidence intervals. These estimate the range of results expected in 95 out of 100 replications (repetitions of the study). As such they are a measure of reliability. The relative size of the dot shows the contribution of the individual finding to the combined overall meta-analysis⁴¹.

Positive effects, right of the central vertical line (> 0.00), show that there was an improvement in the treatment group/intervention condition pupil outcomes compared to the control or comparison period. Negative effects, left of the central vertical line (< 0.00), show that the control group or comparison period performed better. The effect size used in the analysis is r (a non-parametric conversion from d)²⁹, used because many of the teacher studies produced data that was not normally distributed. Readers may be more familiar with Cohen's d (used by John Hattie in his tables³⁰). For comparison, we have included this on the right.

Where we have marked an effect size with an asterisk [*, ** or ***], this shows that the effect size is statistically significant. In this case, we can use the 'p-value' to assess whether there is a significant difference between the intervention and control condition data. For example, $p < .05$ means a smaller than 5 in a 100 probability; $p < .001$ a less than 1 in a 1,000³¹. By convention, researchers use the word 'significant' when p-values cross a threshold (usually $p < .05$). Probability is a function of effect size and sample size. Large effects can be significant with small sample sizes and conversely small effect sizes significant with large samples.

On the left-hand side of the forest plot is a brief description of the intervention, on the right the year group and type of assessment. At the base of the plot there is the pooled effect size and confidence interval across the sub-group analysis

Because the majority of the teacher studies were aiming to assess a 'no-harm' hypothesis (i.e. that there would be no change in attainment associated with reducing workload) non-significant effects are arguably equally important in the interpretation of the findings.

²⁹ Rosenthal (1991)

³⁰ Hattie (2009; 2012). Cohen's d is also the measure that underpins the months' gain calculation in the Education Endowment Foundation's reporting of what works (see Elliot Major and Higgins, 2020).

³¹ In the forest plots in Appendix 2, * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

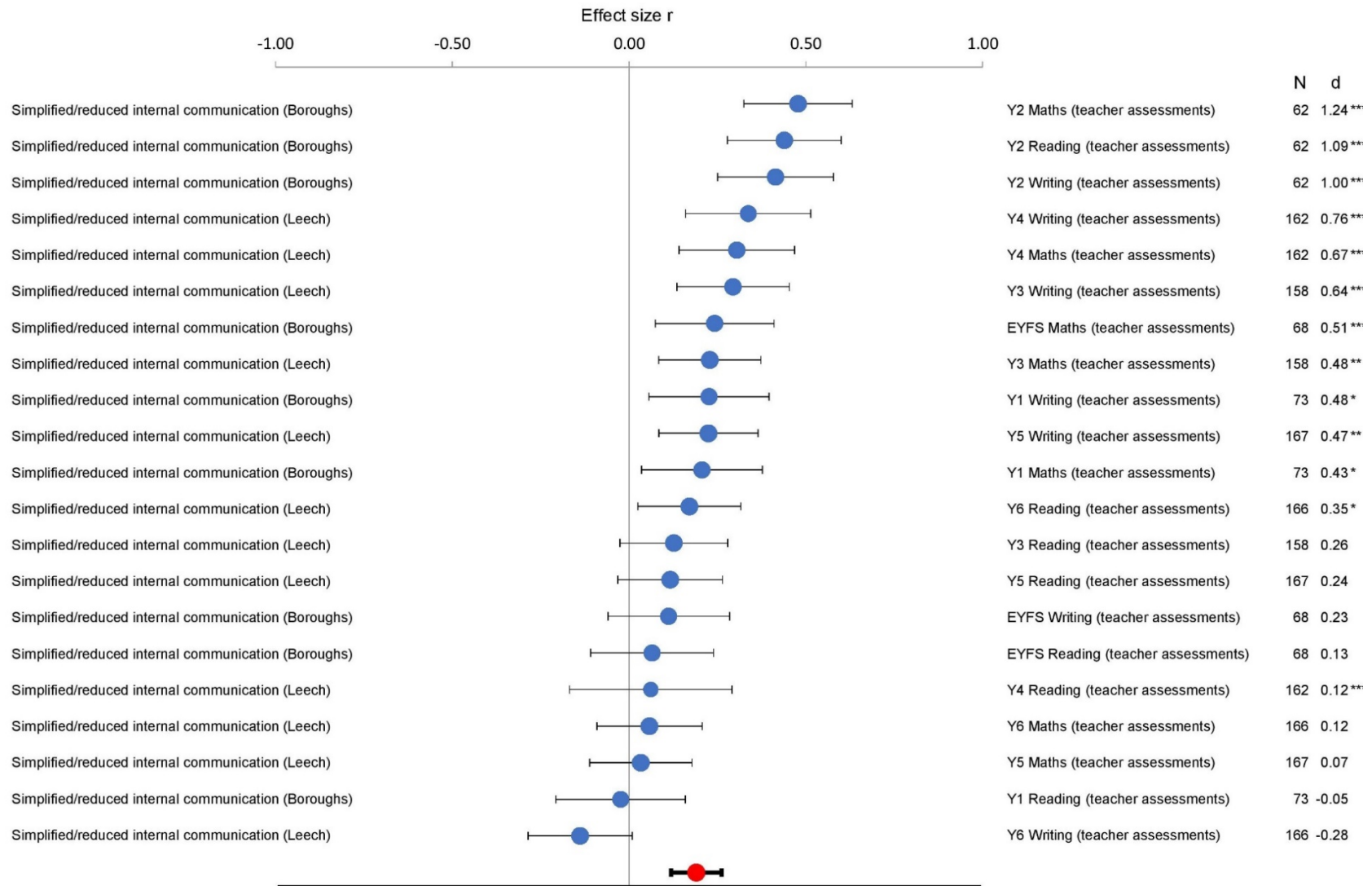


Figure 5: Communication workload reduction interventions (forest plot)

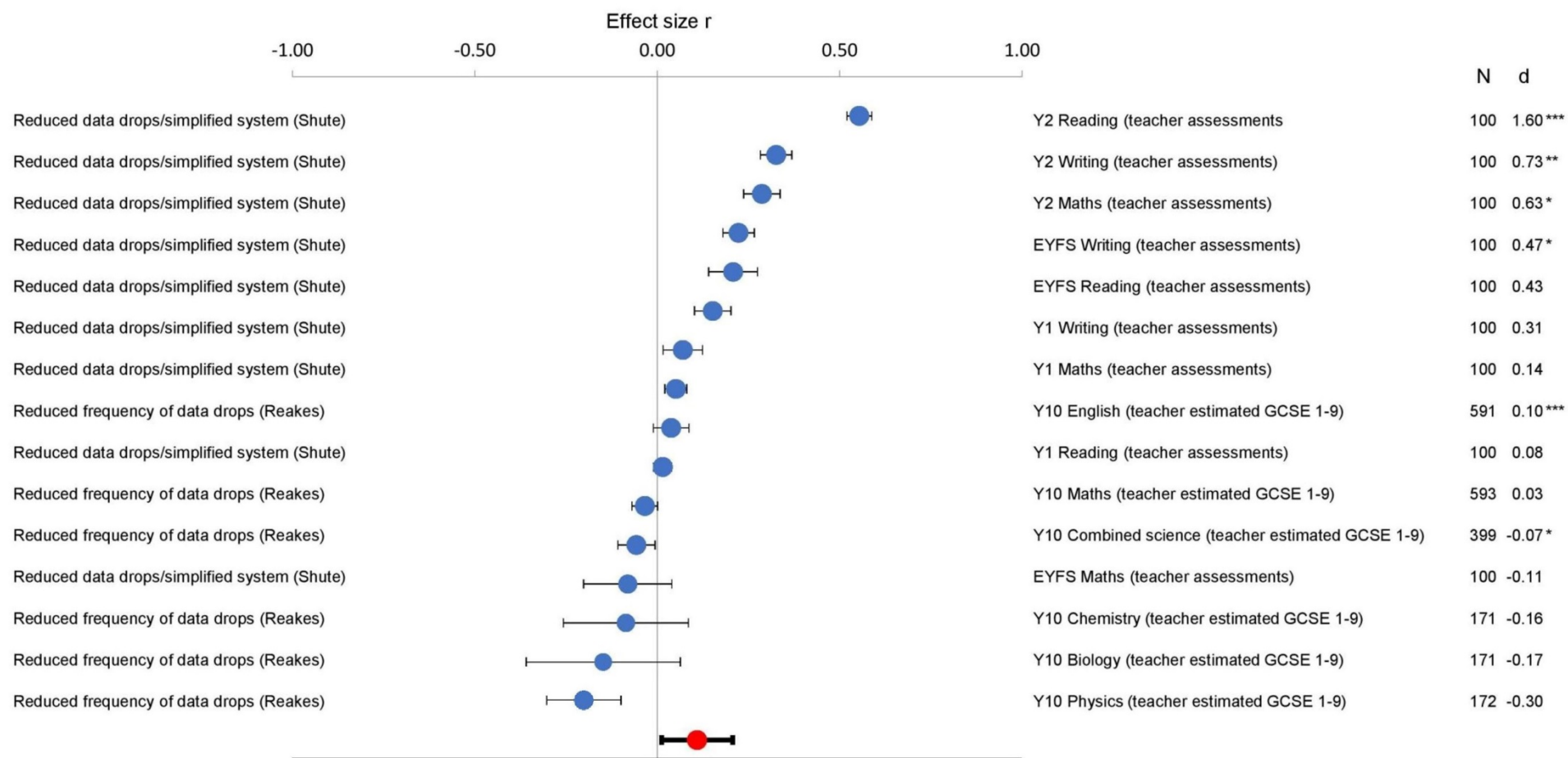


Figure 6: Data recording and reporting workload reduction interventions (forest plot)

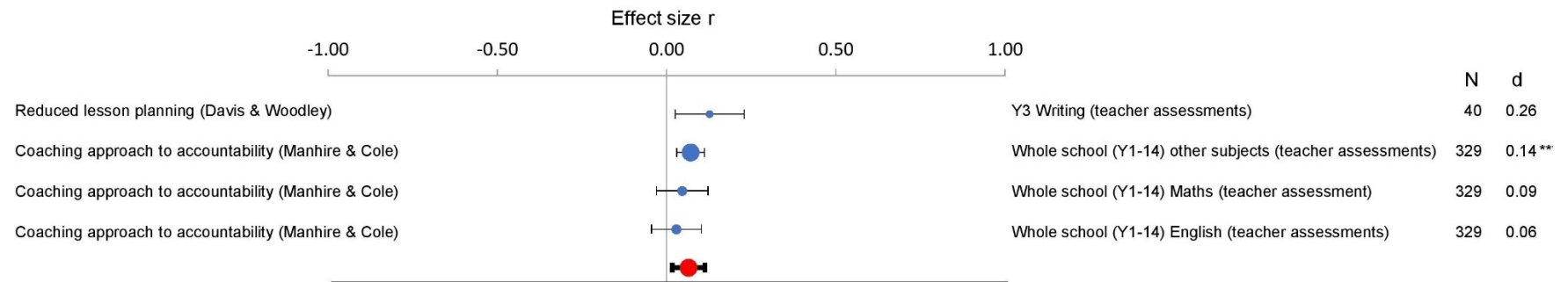


Figure 7: Lesson observation and monitoring workload reduction interventions (forest plot)

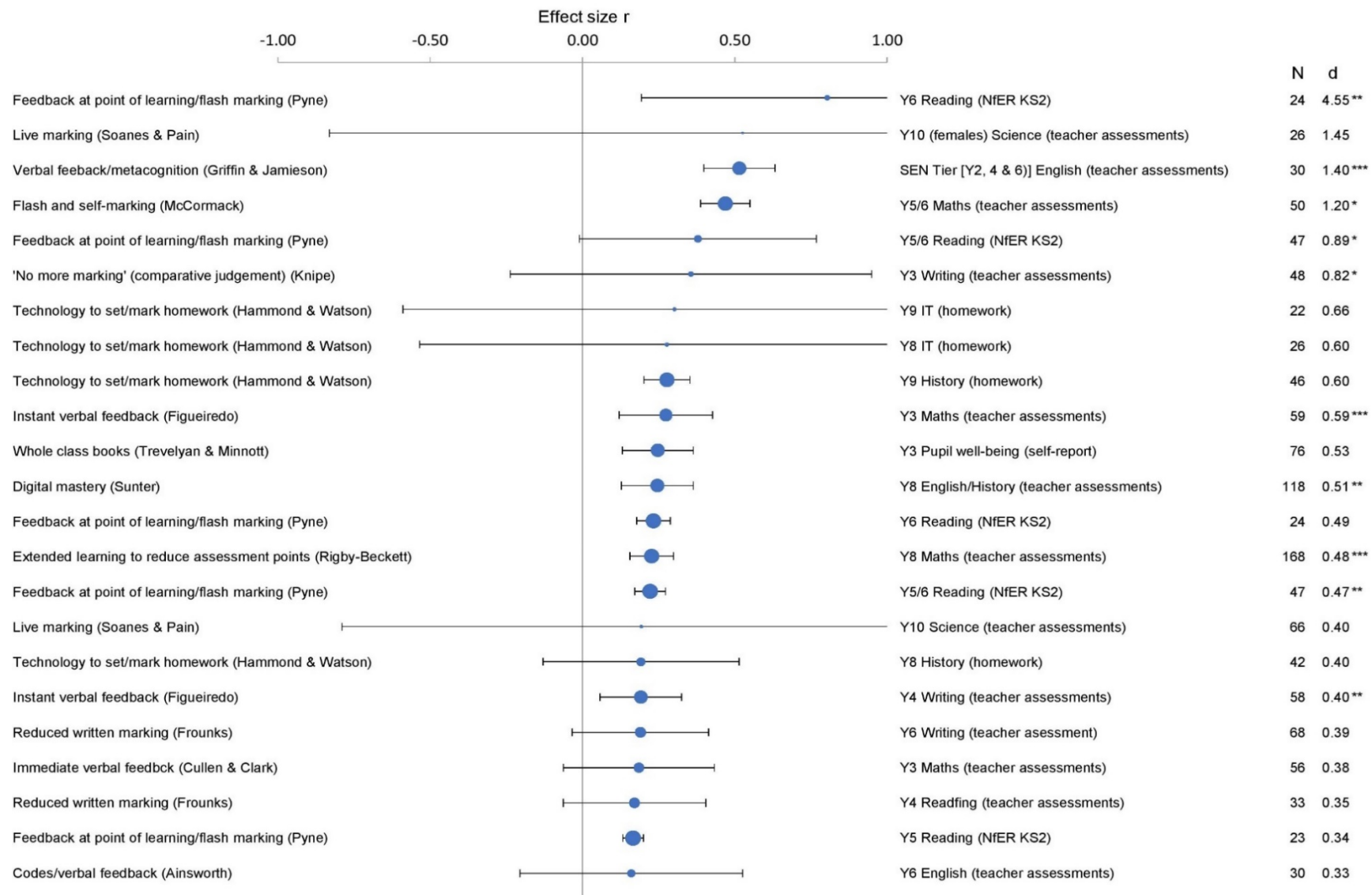


Figure 8: Marking and feedback workload reduction interventions (forest plot) Part 1

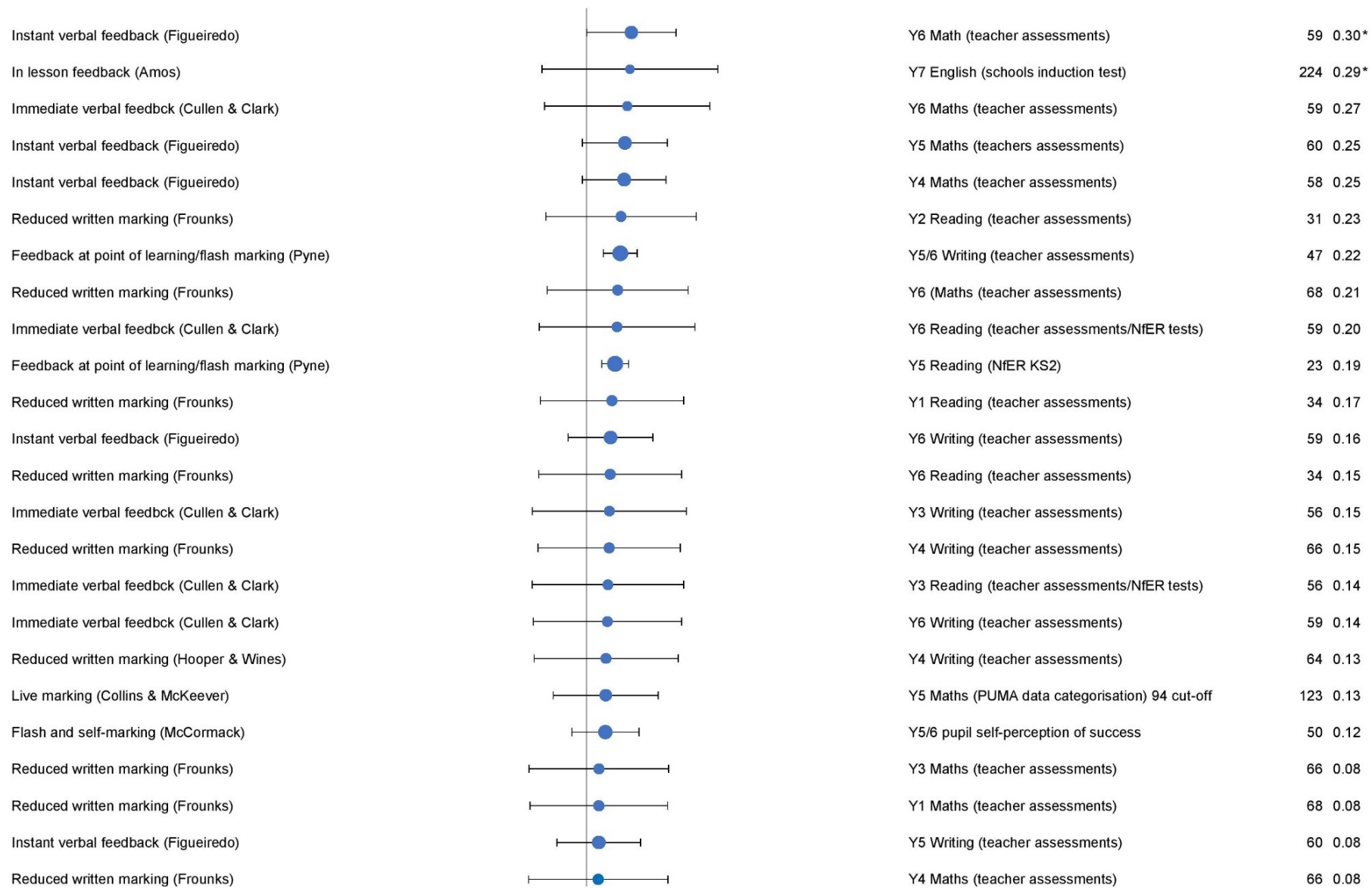


Figure 9: Marking and feedback workload reduction interventions (forest plot) Part 2

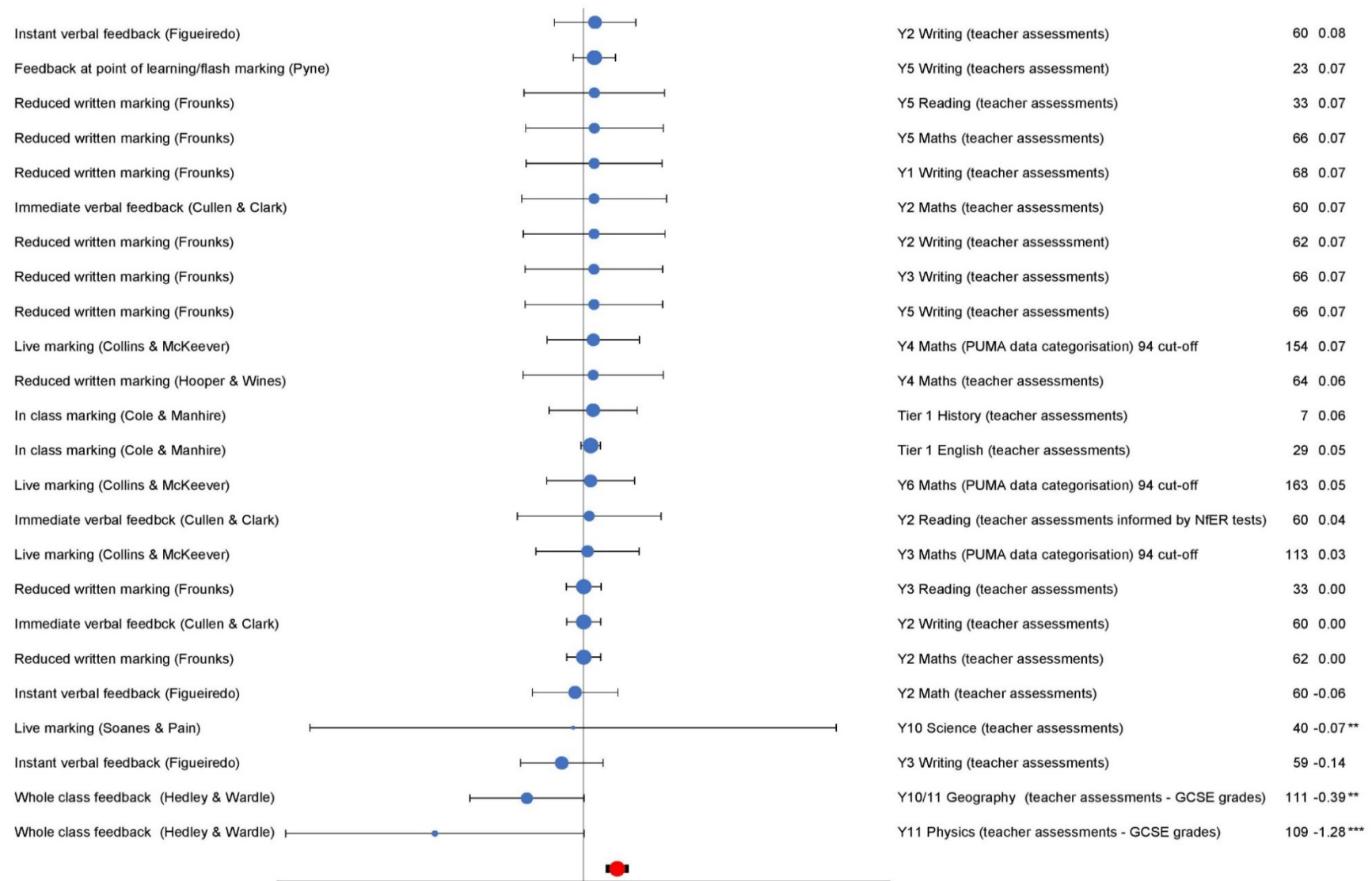


Figure 10: Marking and feedback workload reduction interventions (forest plot) Part 3

Teacher research designs included in the meta-analysis

Teachers chose a wide range of research designs, with most having multiple planned year group or subject level replications. Sample sizes ranged from 7 to 593 with an average sample size of 98.40. Frequently, because the school had already agreed which teachers would be involved in workload reduction (or whole school implementation was desired) random allocation was not possible. In these instances, teachers opted for a form of quasi-experimental design or cohort study.

The following designs produced the following number of pupil outcome effect sizes:

- 15 from randomised controlled trials
 - between-participant designs (independent measures) (7)
 - matched pair designs (7)
 - within-participant designs (repeated measures) (1)
- 21 from non-randomised controlled trials
 - between-participant (parallel group) (3)
 - case-matched (parallel group) (17)
- 68 from retrospective cohort studies
 - controlled (28)
 - uncontrolled (using within-participant data) (40)
- 9 mid-trial results 1 a non-randomised stepped wedge design

© Crown copyright 2020

This publication (not including logos) is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

To view this licence:

visit www.nationalarchives.gov.uk/doc/open-government-licence/version/3

email psi@nationalarchives.gsi.gov.uk

write to Information Policy Team, The National Archives, Kew, London, TW9 4DU

About this publication:

enquiries www.education.gov.uk/contactus

download www.gov.uk/government/publications