Title Page

Experimental trials and 'what works?' in education: The case of grammar for writing.

Dominic Wyse (UCL Institute for Education) and Carole Torgerson (Durham University)

Abstract

The place of evidence to inform educational effectiveness has received increasing attention internationally in the last two decades. An important contribution to evidence-informed policy has been greater attention to experimental trials including randomised controlled trials (RCTs). The aim of this paper is to examine the use of evidence, particularly the use of evidence from experimental trials, to inform national curriculum policy. To do this the teaching of grammar to help pupils' writing was selected as a case. Two well-regarded and influential experimental trials that had a significant effect on policy, and that focused on the effectiveness of grammar teaching to support pupils' writing, are examined in detail. In addition to the analysis of their methodology, the nature of the two trials is also considered in relation to other key studies in the field of grammar teaching for writing and a recently published robust RCT. The paper shows a significant and persistent mismatch between national curriculum policy in England and the robust evidence that is available with regard to the teaching of writing. It is concluded that there is a need for better evidence-informed decisions by policy makers to ensure a national curriculum specification for writing that is more likely to have positive impact on pupils.

Key words: Experimental trials; research evidence; grammar teaching; teaching writing

Experimental trials and 'what works?' in education: The case of grammar for writing.

Dominic Wyse (UCL Institute for Education) and Carole Torgerson (Durham University)

One of the most important questions in education is: what works best to help children and young people learn? Possible answers to this question are a daily reality for teachers and their pupils, and the question is also of great concern to wider society, not least because of governments' significant expenditure on education, and the expectations that arise from this expenditure. Society expects schooling to enhance pupils' learning as a result of teaching that is *effective*.

Over the last decade, across the world, the political impetus to examine 'what works?' as part of educational effectiveness has coincided with the growth in the use of two specific research designs to evaluate educational policy and practice: international comparative surveys using large data sets, and more recently a growth in experiments and quasi-experiments (Connolly, 2015). International comparative work, including the testing of representative samples of pupils, is a prominent feature in education policy evaluation in both low-income and high-income nation states. Examples include: the goal-driven approach of the United Nations *Sustainable Development Goals* (United Nations, 2017); the test-driven comparisons of specific aspects of education such as literacy (UNESCO, 2006); and large scale international surveys such as the Programme for International Student Assessment (PISA, for secondary schooling), the Progress in International Reading and Literacy Study (PIRLS, for primary schooling) and the Trends in International Maths and Science Study (TIMSS, covering both primary and secondary) that combine pupil testing with surveys exploring some aspects of educational policies in the comparator countries. The research designs used in these international surveys are able to establish *correlations* between education policies and outcomes, but are not able to establish whether such policies *cause* the observed outcomes. A causal relationship to demonstrate effectiveness requires a design which features a control group, i.e., a 'true' experiment – a randomised controlled trial (RCT) - or a quasi-experiment (QE). The extent to which studies using a variant of RCT or QE design can establish stronger or weaker causal inference also depends on the robustness *within* the design and its conduct. Although RCTs comparing the curriculum policies of whole countries are not feasible, RCTs of specific approaches to teaching are feasible, not least in areas such as literacy that are included as a comparator in most international analyses of the kinds described above.

A paramount source of information about effective teaching should be research; however, the extent to which *education* research has contributed answers to the questions of teaching efficacy and effectiveness is fiercely debated. As early as 1972, in the United States (US) congress there was a view that education research was "mediocre and useless" (Kaestle, 1993, p. 27). Thirty years later, it was observed that education in the US had been

dragged “kicking and screaming, into the 20th century” (Slavin, 2002, p. 15) as a result of developments of education policy linked to “scientifically based research”, such as the Elementary and Secondary Education Act *No Child Left Behind,* and emphasis on “proven, comprehensive reform models” (*op. cit.)*. At the time, the US Office of Educational Research and Improvement invited nomination of programmes to be evaluated, ultimately using experimental designs, by third-party evaluators (*op. cit.*).

In the UK, the debate about the capacity of education research to contribute to questions about effectiveness was reignited around 20 years ago. A trend of criticisms of education research was typified by the *Teacher Training Agency* Annual Lecture in 1996 given by David H. Hargreaves who was, at the time, a Professor of Education at the University of Cambridge. Hargreaves’ strong criticism of education research included his opinion that it was poor value for money in relation to improving education in schools, and that the teaching profession had been inadequately served by education research. In a comparison with medicine Hargreaves’ conclusion was that, “In education we too need evidence about what works with whom under what conditions and with what effects” (Hargreaves, 1996, p. 8). More recently, the debate came to prominence as a result of the work of the medical doctor, research fellow and journalist Ben Goldacre (Goldacre, 2013). One notable aspect of Goldacre’s 2013 argument is how similar it was to some of the points made 20 years previously by Hargreaves. For example, the advocacy for RCT designs, the idea that research evidence about education practice is weak, and comparisons with medicine were all addressed in Hargreaves’ original lecture.

The aim of this paper is to examine the use of experimental trials in relation to evidence about effective teaching, and to consider some links between research and national curriculum policy. To do this we selected one important research area as a case: the teaching of grammar to help pupils’ writing. The teaching of grammar for writing is a useful case because the topic has attracted a fierce ideological debate as well as a significant number of experimental and quasi-experimental trials evaluating interventions to improve writing. Unlike previous work that has had a main focus on the methodology of experimental trials *or* on the implications of evidence from experimental trials for an aspect of policy and/or practice, our argument is built on an in-depth analysis of methodology *and* research outcomes in a specified aspect of education, namely the teaching of grammar to improve writing. Through examination of methodology and a substantive topic a stronger case can be made in relation to which teaching method is likely to be effective.

The paper begins with a historical account of the debate about research evidence and experimental trials in education. We then review the research evidence on grammar teaching for writing. The curriculum policy context for grammar teaching is seen in our brief description of representations of grammar in national curricula internationally, and an account of the development of England’s national curriculum of 2014. The main part of the paper is a detailed exploration of two well-regarded experimental trials, published in peer-reviewed research journals, and focussed on evaluating the

role of grammar teaching in supporting the development of writing. The studies examined similar approaches to teaching grammar in the same phase of education, and both papers had a recognised impact on policy and practice. The studies were also chosen because, although they addressed very similar teaching approaches, they came to different conclusions about their effectiveness. One of the two papers concluded that grammar teaching to support writing was *not* effective, whilst the other paper concluded that it *was* effective. The important considerations for the argument in the present paper are: a) what the comparison of the two studies reveals about the methodology of experimental trials; b) the extent to which the outcomes of either of the two studies are replicated in other experimental trials in the same field; and c) as a result of considering a) and b) whether the research evidence of grammar for writing is appropriately reflected in national curriculum policy in England, and what the implications are for research, policy and practice.

Experimental trials in education research

Although the RCT is widely used in medical research, one of its first known uses to investigate human activity (as opposed to RCT use in the natural sciences) in the modern period was in the field of education. In the early 1930s in the US, Walters undertook two randomised experiments in the field of education (1931; 1932). In a university setting Walters randomised the selection of members of the freshmen class in the School of Mechanical Engineering in Purdue University. Some of the freshmen were allocated to mentoring delivered by five seniors with a "good scholarship record, pleasing personality, excellent health and fine social environment", and some were allocated to a control mentoring condition. Academic outcomes were then measured, and Walters concluded that the students in the mentoring condition had better outcomes than the students in the control condition (no mentoring). Walters' experiment is the first known use of the term 'random sampling' – or randomisation – to form equivalent groups: "The 220 delinquent freshmen were divided into two groups by *random sampling."* The following year Walters undertook a replication trial with a much larger sample size and random allocation to one of 3 'arms' – mentoring by seniors, mentoring by Faculty members, and a control condition (*op. cit.*) and he concluded that the senior students were more effective in personal mentoring in reducing drop-out or exam failure than the Faculty members.

Between 1900 and the 1960s many 'explanatory' experiments were undertaken in the field of education, sometimes using randomisation. These tended to be conducted by educational psychologists, working in psychology laboratories, investigating basic psychological processes relevant to learning. Between the 1930s and 1970s many RCTs in education were undertaken in the US (some large scale), but there was a dearth of high quality RCTs in education research in UK. Between the 1970s to the 2000s, there were very few large scale RCTs in education in US, as the design had largely fallen out of favour, although there were a few notable exceptions.

After the 50 year lull in activity, greater emphasis on experimental trials to inform education policy in the US and the United Kingdom (UK) became evident. This step-change in the history of the use of the design was largely driven by two distinct policy initiatives on either side of the Atlantic. In 2002, when George Bush enacted the *No Child Left Behind* Act, the subsequent creation of the Institute of Education Sciences (IES) led to public investment in the use of experimental design to evaluate education policies and interventions. The legislation mandated the RCT as the design of choice for evaluating education interventions: "Scientifically valid educational evaluation employs experimental designs *using random assignment*, when feasible, and other research methodologies that allow for the strongest possible causal inferences when random assignment is not feasible" (p.5). Since that time, over 200 experiments and quasi-experiments have been funded by the IES and undertaken in education in the US. The IES does fund quasi-experiments, but only if randomisation is not thought to be feasible, which occurs in rare circumstances, where, for example, it may be deemed unethical to undertake random allocation.

The UK equivalent to the greater use of experimental trials in education in the US was the creation of the Education Endowment Foundation (EEF) in 2011, and its requirement that "...all EEF projects will be rigorously evaluated by independent experts in educational research according to minimum standards … The impact of projects on attainment will be evaluated, where possible, using randomised controlled trials" (EEF, 2017, online). The EEF has now funded over 120 RCTs and quasi-experiments, evaluating education policies and practices; and, similar to the IES policy, only funds quasi-experiments where randomisation is not feasible.

The similarity of the criticisms of educational research made by both Hargreaves and Goldacre, alluded to earlier in this paper, seemed to indicate that little had changed in relation to the nature of educational research; however, the evidence shows a different picture. Between 1980 and 2015 the number of RCTs in education demonstrated significant increases, particularly from 2006 onwards (Connolly, 2015). According to Connolly's analysis, although the US and Canada are still responsible for undertaking the majority of RCTs (approximately 375), the UK had a significant number (approximately 80), in comparison to much larger population areas (e.g., rest of Europe approximately 140; Australia/New Zealand 50) (*op. cit.*). More than 200 of the RCTs have focused on interventions taking place over a full academic year or longer (short duration RCTs was a criticism made by Slavin, 2002). Approximately 540 RCTs focused on: physical health and wellbeing; behaviour and social wellbeing; and professional training. Approximately 90 evaluated literacy/English language interventions; and approximately 225 focused on other academic interventions and outcomes, study-related skills, and numeracy/maths interventions (*op. cit.*). There is less evidence about the frequency of use of other experimental designs.

During this period of growing emphasis on RCTs in education the longstanding philosophical critique of 'positivist' methodologies also continued. RCTs, as part of evidence-based education research, have been

criticised because they are reductionist and not appropriate for the evaluation of educational interventions which, as a result of the complexity of the social context, are necessarily more challenging compared with experiments in the natural sciences (e.g., Morrison, 2001). But others have countered with the opinion that, for some research questions, a well conducted RCT is the strongest research design when seeking to compare *effectiveness* of interventions. For example, the potential of RCTs was seen in a complex intervention on sex education at secondary education level that paid careful attention to the methodological challenges of evaluation in the real world of secondary schools (Moore, Graham and Diamond, 2003). Another more recent strand of the debate has linked a critique of positivism with support for 'realist' approaches (e.g., as recommended by the criminologists Pawson & Tilley, 1997), including the important idea that 'what works' should have a central focus on who an intervention works for, and the context in which a specific intervention can work. In an exploration of Pawson and Tilley's ideas, Bonell et al (2012) acknowledge the importance of attention to theories of causal mechanisms but critique the realist position on the grounds of: misunderstanding of the use of counterfactuals; the resultant limit on findings based on plausibility rather than on probability (in a statistical sense); and on a lack of acknowledgement that well-conducted experiments do include attention to mechanisms and context but also are able to assess causal attribution, something which realist approaches cannot do. Stronger experimental studies have, for some time, recognised context and methodological limitations. This recognition is evident in the seminal book on experimental design: "The experiment is not a clear window that reveals nature directly to us. To the contrary, experiments yield hypothetical and fallible knowledge that is often dependent on context and imbued with many unstated theoretical assumptions … In this sense, all scientists are epistemological constructivists and relativists, the difference is whether they are strong or weak relativists." (Shadish, Cook & Campbell, 2002, p. 29). More recently, the epistemological debate has also been informed by ongoing developments in mixed methods design and methodology including the more routine use of process evaluation, or embedded ethnography, as part of RCTs. These developments include the recognition that the dualisms and intellectual tensions that are part of mixed methods methodology, and of understanding what works, are usefully framed by philosophical pragmatism (Johnson, Onwuegbuzie, de Waal, Stefurak and Hildebrand, 2017).

The teaching of grammar in national curricula

Recent growth of interest in grammar for writing has been clearly evident in developments in national curricula in a range of countries with English as a main language. For example in the *Australian Curriculum*'s English learning area, the language strand is positioned first in the curriculum structure before the strands for literature and literacy. For children aged 10 to 11 this strand includes explicit attention to "sentences and clause-level grammar" and to "noun groups/phrases" and "adjective groups/phrases" (Australian Curriculum, Assessment and Reporting Authority, 2017, online). In the US the *Common Core State Standards* text for English Language Arts for the same age of children specifies reading, writing, speaking and listening, then Language. As

part of the language specification “Conventions of Standard English grammar and usage” (including forming perfect verb tenses; explaining the function of prepositions; etc.) is listed before “Knowledge of Language” and “Vocabulary Acquisition and Use” (National Governors Association Center for Best Practices Council of Chief State School Officers, 2010, online). These kinds of emphases on grammar are not only evident in high-income nations and states but also in other post-colonial countries with historic links to the British Empire, for example in the countries of Africa (e.g., see Wyse et al, 2014).

The emphases in New Zealand’s national curriculum appear to have some differences from the countries surveyed in this paper so far. In *The New Zealand Curriculum* the focus on language is a holistic one, with an emphasis on the making and creating of meaning (New Zealand Ministry of Education, 2007, p. 18). This holistic attention to language is also reflected in the strong place of the indigenous language Te Reo Māori and New Zealand Sign Language, and in the title “an English medium curriculum”. The emphasis on grammar also appears to be different. For example, the specification of “Language features” as part of “Speaking, Writing and Presenting” is positioned last in the list of curriculum requirements, and emphasises the way that pupils should understand grammar as follows: “Use a wide range of text conventions, including grammatical and spelling conventions, appropriately, effectively, and with accuracy” (New Zealand Curriculum, Years and Curriculum Levels, Level Six English).

In the different countries of the UK the national curricula for language and English have differed markedly since political devolution of powers, with England having increasingly more emphasis on discreet elements such as grammar and phonics (Wyse et al., 2013). The importance attributed to grammar by policy makers in England since 2011 can be seen in the intensification of the teaching of formal grammar as part of the subject of English in England’s national curriculum. In the national curriculum of 2014 the programmes of study for writing for nine-year-old to eleven-year-old pupils include statutory requirements for the teaching of “Writing – transcription”, including spelling, handwriting and presentation. These sections are followed by writing composition (planning and drafting), then vocabulary, grammar and punctuation. Increased attention to vocabulary, grammar and punctuation is added through an appendix that includes an emphasis on “explicit knowledge of grammar” (DfE, 2013, p. 75) where pupils in year 3 (seven- to eight-years-old) are expected to understand terminology that includes “subordinate clause”, and for year 6 (ten to eleven-years-old) the need to be introduced, for example, to the “use of the *passive* to affect the presentation of information in a *sentence* (op cit. p. 79, emphasis in original).

In addition to the emphasis in the national curriculum programmes of study, the national statutory tests for 11-year-old pupils in England included for the first time in 2011 a separate spelling, punctuation and grammar test where formal grammar was further emphasised. In addition, the requirements for teacher assessment of writing included a strong emphasis on grammar as part of the assessment criteria. In 2016 these emphases were still in place. For example the national statutory test for Spelling, Punctuation, and

Grammar included a strong emphasis on formal grammar including questions that required knowledge of grammatical terminology (for example, "**27.** Underline the **subordinate clause** in each sentence below." UK Government, 2016, p. 17, emphasis in original). All questions in the paper attracted one mark each. Although the 2016 criteria for statutory teacher assessment of writing, produced by pupils in lessons, included aspects such as "creating atmosphere" in their writing, there was a strong emphasis on usage according to areas of formal grammar such as "passive and modal verbs" and "adverbs, preposition phrases and expanded noun phrases", etc. (Standards and Testing Agency, 2015).

The politics and policies that led to the emphasis on formal grammar in England's national curriculum implemented from 2014 onwards began with a government white paper in 2010 that included the commitment to "Review and reform the National Curriculum so that it becomes a benchmark outlining the knowledge and concepts pupils should be expected to master to take their place as educated members of society" (Department for Education, 2010, p.41). The link between statutory assessment, the curriculum, and school accountability was also made clear: "The National Curriculum will continue to inform the design and content of assessment at the end of key stage two, which will apply to every child and which will provide a guide to the performance of primary schools" (op. cit, p. 42). After publication of the white paper the government commissioned a review of assessment in England led by Lord Bew. Bew's final report noted that "there are some elements of writing – spelling, grammar, punctuation, vocabulary – where there are clear 'right' and 'wrong' answers, which lend themselves to externally-marked testing … Internationally a number of jurisdictions conduct externally-marked tests of spelling, punctuation and grammar … These are essential skills and **we recommend that externally-marked tests of spelling, punctuation, grammar and vocabulary should be developed**." (Bew, P. (2011). p. 60. Bold font in original).

A public consultation on the proposals for the new national curriculum was held between February and April 2013. It attracted 17,312 respondents with 4,576 described as 'non-campaign respondents', and 12,736 described as 'campaign respondents' (i.e., organisations devoted to a particular issue. The report of the consultation made clear that campaign responses were not included in the percentages of answers to questions but were reflected in the commentaries about the answers). 3,682 respondents addressed the question "Do you have any comments on the content set out in the draft programmes of study?" With regard to the teaching of the subject English, and the teaching of grammar within that subject, "There was recognition that the teaching of phonics, punctuation, spelling and grammar was necessary, but some felt that there was an over-emphasis on these aspects." (Department for Education, 2013b, p. 7). It is disappointing that the number of respondents who replied about grammar was not specified in the report as this would have provided some further evidence relevant to the strength of opinion on this issue.

There was also a follow up consultation, open from July to August 2013, on the draft legislative order, which attracted further comment about English and

grammar. Although 21 respondents (11%) supported the greater focus on spelling, grammar and punctuation,

> a total of 36 respondents (19%) however expressed concern in relation to the more demanding grammatical content included for years 2 and 4 … 52 respondents (28%) said the English primary curriculum was too prescriptive, in particular in reference to the level of specification in the appendices [where the grammatical knowledge to be learned by pupils is specified]. These respondents argued that this undermined the aims of the new national curriculum in relation to greater professional freedom and were concerned that this may have implications for the provision of a balanced and broadly based school curriculum. (Department for Education, 2013c, p. 6)

One interpretation of these data in the second consultation is that 47% of respondents were critical of the grammar specified in the national curriculum and its appendices, but 11% thought the emphasis on correct use of Standard English was commendable. An overall negative response to the proposed attention to grammar did not result in changes to this element of the national curriculum.

Reservations about the nature of the specifications for grammar teaching in the national curriculum and its associated statutory testing continued to cause disagreement. The main government advisor for grammar in the statutory assessment system described the process of determining the curriculum for grammar as “chaotic” and that “We started off with the primary curriculum, which we were a bit unconfident about as none of us had much experience of primary education” (Mansell, 2017). In April 2017 A House of Commons Education Select Committee report on assessment in primary schools concluded that:

> One issue with the writing assessment is the focus on technical aspects, like grammar and spelling, over creativity and composition. We are not convinced that this leads directly to improved writing and urge the Government to reconsider this balance and make spelling, punctuation and grammar tests non-statutory at Key Stage 2 (House of Commons Education Committee. (2017). p. 3)

This brief account of some of the work that led to greater emphasis on grammar in England’s national curriculum, and subsequent implications, shows that research evidence, of any kind, had insufficient consideration and influence on the national curriculum of 2014. Further corroboration of problems with attention to research evidence was detailed by BERA President Mary James (BERA, 2012), one of the expert group advising on the national curriculum. In addition, reflecting on his time as a Minister for Schools under Secretary of State for Education Michael Gove, David Laws claims that decisions were made “not based on evidence but on hunch” (Wilby, 2017, online) and that Gove had a particular weakness for basing decisions on “ideology and personal experience” (op. cit.).

Research evidence on grammar for writing

The place of grammar in education has been a point of debate for at least 200 years, in part because it has been repeatedly linked with the development of the concept of 'standard' English (Crystal, 2004). In the 21st century general interest in grammar teaching as an element in the teaching of writing continued (Wyse, 2001; Andrews et al, 2004a; Andrews et al, 2004b; Myhill & Watson, 2014). In 2001, as a result of a comprehensive narrative review of empirical studies, it was concluded that:

> The findings from international research clearly indicate that the teaching of grammar (using a range of models) has negligible positive effects on improving secondary pupils' writing. Of further concern is the negative impact on pupils' motivation. In the [National Literacy Strategy] Framework for Teaching the move towards the teaching of grammatical 'technical vocabulary' such as adjective; noun: collective, common, proper; pronoun: personal, possessive; verb, and verb tense to six and seven year-old children in England is highly questionable. It is regrettable that there is not more evidence about primary pupils; however, *the developmental arguments that such teaching is inappropriate at primary level are persuasive.* (Wyse, 2001, p. 422, emphasis added)

This finding was subsequently supported in two systematic reviews (SRs) undertaken by one of the authors of this paper and colleagues (Andrews et al 2004a; 2004b). In the first systematic review evaluating the effect of grammar teaching (syntax) in English on 5-16 year-olds' accuracy and quality in written composition, Andrews et al (2004a) concluded there was insufficient high quality evidence to "counter the prevailing belief that the teaching of the principles underlying and informing word order or 'syntax' has virtually no influence on the writing quality or accuracy of 5 to 16 year-olds" (Andrews, 2004a). This conclusion applied to both the 'traditional' approach of emphasising word order and parts of speech and the 'transformational' approach, based on transformational-generative grammar. The current picture of robust research in relation to grammar teaching to support pupil's writing is shown in Tables 1 and 2.

Insert near here:
Table 1 A selection of key meta-analyses, and influential single experimental studies (primary/elementary education).
Table 2: A selection of key meta-analyses, and influential single experimental studies (secondary education).

As the evidence summarised in tables 1 and 2 shows, as far as primary/elementary education is concerned there is strong evidence that grammar teaching of a range of types, but particularly traditional grammar teaching, is *not* effective for improving pupils' writing. There is evidence that sentence-combining is effective but no experimental studies have been carried out in the UK. At secondary education level there is a slightly more mixed picture. The majority of the evidence suggests that, apart from

sentence-combining, grammar teaching is not effective for improving pupils' writing. However one robust study, Myhill et al. (2011), showed that contextualised grammar teaching was effective for improving secondary pupils' writing, although the approach was more effective for higher attaining pupils.

In about 2010, a challenge to the longstanding view, that grammar teaching was not the most effective way to improve writing, emerged from researchers in the UK. For example, in an interview, it was stated: "… what we have for the first time ever, internationally, is research evidence that shows that the teaching of grammar can have an impact on children's writing skills. But the way that we taught it was completely unique" (Education Arena, ND, online). More specifically, it was claimed in relation to a RCT evaluating grammar teaching, that, "the strong positive effect of the intervention signals for the first time the potentiality of grammar as an enabling element in writing development and evidences a clearly theorised role for grammar in writing pedagogy" (Myhill et al. 2011, p. 162). The overall claim that an experiment had demonstrated that grammar teaching could have a positive impact on secondary pupil's writing skills (Myhill, et al., 2011)[1] was in opposition to the conclusions of a seminal experiment published in 1976 that had evaluated a similar grammar intervention to improve writing (Elley, et al. 1976). This experiment had concluded that grammar teaching did *not* have a positive effect on writing.

The Elley and Myhill studies have been selected for detailed comparison in this paper because both were experimental trials (one a RCT, the other a QED), both were regarded as having significant wider impact, including political and professional impact, and both were published in peer-reviewed research journals. Elley et al's (1976) quasi-experiment has been regarded as one of the most rigorous in the field, having been reprinted by the US National Council for Teachers of English (NCTE) because it was "so important [and] a model of evaluation" (Elley et al., 1976, p.5). The impact of Myhill et al's (2010) randomised experiment was recognised by the UK Economic and Social Research Council because it "shaped policy and curriculum development in England - including the first author leading the advisory group of four writing the Grammar Annex of the Primary English curriculum; participation in the KS2 English Test team; and providing expert testimony in discussions of the English curriculum revision with the Minister of State for Schools (2012) … Professor Myhill also provided evidence for the new Australian curriculum" (ESRC, 2016, online). As will be demonstrated in detail below, each of these studies had relative strengths and limitations, not least in their basic design; however, due to the availability of any other experimental research addressing the same teaching approaches, and having had the same reach and significance as these two studies, we consider such a comparison relevant. We do acknowledge, however, the challenges and limitations in making the comparison, given the differences between the two

[1] For brevity, in the rest of the paper we refer to this study as the 'Myhill study' and the Elley et al study as the 'Elley study'.

studies, in particular, in terms of the countries and years in which they were undertaken and published.

Below we discuss the two studies in detail, in terms of the intervention and control conditions, design and features and components of design, and assess the main methodological strengths and limitations.

Teaching methods for the control and intervention groups

An important consideration for any experimental trial, or a systematic review of trials, (and for our comparison in this paper) is that the nature of the teaching methods are clearly specified in the publication, and are a suitable comparison, including a comparison with at least one appropriate control group. In both the Elley and Myhill studies a form of contextualised teaching of grammar was one of the interventions evaluated.[2]

For the Elley study one of the intervention groups used an approach called The Transformational Grammar Course (TG), based on Jerome Bruner's concept of the spiral curriculum. In this intervention group, all the activities "were related to the central core of each strand of the curriculum, thus giving it [the teaching approach] a clear and consistent unity of purpose." (p. 8) The TG intervention included the three strands of a) Grammar (Transformational); b) Rhetoric; and c) Literature.

One control group in the Elley study used an approach called 'Reading-Writing', which included rhetoric and literature (as did the TG intervention) but substituted extra reading and creative writing instead of transformational grammar. The other control group used an approach called 'Let's Learn English': a traditional approach to grammar including the learning of parts of speech and some applications of them.

The Elley intervention and control groups "had approximately 574 periods of English in the three years, distributed such that each class had similar proportions of morning and afternoon periods, and of time spent on literature, on composition work, and evaluation exercises." (p. 10). Although it was claimed that "no detectable bias was apparent in their approach to their teaching of any of the [grammar] courses" (p. 10), there was no attempt to establish fidelity to the interventions, which is a significant limitation of this study.

In the Myhill study, the intervention took place over three weeks per term for one school year: "for both the intervention and comparison groups, the learning focus, the period of study, the learning objectives and the assessed

[2] Contrary to Myhill et al's claim that the Elley study did *not* include contextualised grammar teaching as one of the interventions it is evident from the description in the Elley paper of the Transformational Grammar (TG) approach, inspired by Bruner's spiral curriculum as we show below, that it did (also confirmed in a personal communication with Warwick Elley in 2013).

written outcomes were the same" (p. 147). For the intervention group the teaching designed by the project team "explicitly sought to introduce grammatical constructions and terminology at a point in the teaching sequence which was relevant to the genre being studied (p. 148). The intervention and control groups were both taught the same writing genre over a three-week period once per term of the year of study. The teaching in both groups also addressed the same learning objectives from England's national framework for English that was being implemented at the time. The intervention in the Myhill study "comprised detailed teaching schemes of work in which grammar was embedded where a meaningful connection could be made between the grammar point and writing." (p. 146) The Myhill intervention was based on the following principles:

- The grammatical meta-language is used but it is always explained through examples and patterns.
- Links are always made between the feature introduced and how it might enhance the writing being tackled.
- The use of **'**imitation**'**: offering model patterns for students to play with and then use in their own writing.
- The inclusion of activities which encourage talking about language and effects.
- The use of authentic examples from authentic texts.
- The use of activities which support students in making choices and being designers of writing.
- The encouragement of language play, experimentation and games. (Myhill, et al, p. 148)

There are two issues with the specification of teaching approaches in the intervention and control groups in the Myhill study. Firstly, in each term of delivery both intervention and control groups experienced teaching where "… using grammar accurately and appropriately…" was a pre-planned objective in the scheme of work. In the intervention groups: "The intervention comprised detailed teaching schemes of work in which grammar was embedded where a meaningful connection could be made between the grammar point and writing" (p. 7). The control groups *did* receive some grammar teaching, as the teaching objectives used by both intervention and control groups specify: "Autumn Term/Narrative Fiction/Using grammar accurately and appropriately" (p. 7 emphasis added). Secondly, like the Elley study there were no checks for fidelity in either condition: "Fidelity is a problematic concept in a naturalistic educational setting such as this, as identical implementation of the intervention teaching materials is neither possible nor desirable. Teachers [in the intervention] were not asked to follow the lesson plans rigidly; they were allowed to adapt materials to suit the needs of their students, but were also asked to remain as close as possible to the materials." (p. 9). So it is possible that the grammar teaching delivered by the teachers in the control condition included contextualised teaching of the Myhill kind; or that they used formal grammar; or more probably that there was a mixture of approaches. As a result the specific role of grammar was not isolated in the trial. It cannot be definitively claimed that it was the grammar that was effective, or not effective, in either of the Myhill or the Elley studies because it could have been a range

of factors, including simply better teaching as a result of the training, i.e., the Hawthorne effect.

Site, sampling, design, and allocation to groups

The Elley study took place in one large co-educational high school on the outskirts of Auckland city. At the start it involved 248 pupils in eight matched classes of average ability who were taught, observed and regularly assessed from the beginning of third-form year in February 1970 to the latter part of the fifth-form year in November 1972. The results of the reading test, of the assessment of the distribution of fathers' incomes, the secondary certificate of education exam results, and the inclusion of 15% Polynesian pupils indicated a so-called 'normal' sample. Elley noted that, "At the outset, one bright and three slow-learning classes were deliberately excluded from the total third-form intake of 380 pupils, thus rendering it more homogeneous, and increasing the chance of identifying systematic differences between groups" (p. 7). The experimental pupils "were classified into eight matched classes of 31 pupils" on the basis of a number of tests, and additional matching criteria were "ethnic group, sex, contributing school, and subject options" (ibid). Although the pupils were allocated as individuals to the eight classes, the study – after this allocation – works as a cluster trial as the pupils in the eight classes were taught together. The three experimental groups contained 3, 3 and 2 classes respectively, and the pupils were tested during the intervention period and at the end.

Limitations of the sampling and grouping in the Elley study include: the lack of random allocation to groups; the small sample size of 8 classes or clusters in total split between 3 groups (statistical methodologists state that, as a minimum, there should be 4 clusters per group in a cluster randomised trial Donner & Klar, 2000); and the fact that it was undertaken in only one school, thereby reducing external validity. This latter issue introduces the possibility of potential 'contamination' or 'spill over' of the intervention and control conditions between the groups, and whether this occurred or not is not clear. The lack of random allocation is important because random allocation minimises any selection bias at the start of the experiment. In a quasi-experiment matching is sometimes used to ensure baseline equivalence, as in this case. The classes were matched on a number of variables including performance on a number of pre-tests. However, Elley et al did not report the results of the matching and, therefore, we have to take on trust that the classes were, in fact, matched on the observed variables. Also, matching cannot account for imbalance on unknown variables which can in turn introduce a potential source of bias which could affect outcome. Furthermore, Elley and colleagues did not adjust for the clustering in their analysis and instead analysed their data as though this was an individually allocated quasi-experiment; and, although they made some attempts to control for teacher effect, given the small sample size (see above), this would not have been possible. This study also suffered from high attrition of pupils – over 30% by the final follow-up in Year 3.

In the Myhill study the authors identified a sample of 32 mixed comprehensive schools from the South West and Midlands areas of England. Lists of schools from local authorities were randomly sampled until the desired sample size was achieved. Once the schools had been recruited, a year 8 class was selected (with children aged 12-13 years) and the classes were stratified according to the teachers' 'Grammar Subject Knowledge' (GSK) then the classes were randomised using a random number generator. In these respects, the Myhill study is of higher design quality than the Elley study: a random sample of schools in two geographical areas in the UK was used to form the intervention and control groups, thereby increasing external validity. The design was a large cluster randomised controlled trial, with school as the cluster, thereby minimising the potential for contamination between groups.

Tests and measures

Data for the Elley study were collected in the form of a series of set essays at the end of each year marked by teachers from neighbouring schools plus a battery of standardised tests. The essays were assessed by carefully-briefed panels of English teachers from neighbouring secondary schools. In the first year each pupil wrote four essays which were assessed by four markers, working independently using a 16-point scale that included criteria for content, organisation, style and mechanics. In subsequent years the number of essays was reduced to three essays and two markers, apparently with no loss of reliability. The battery of tests included: 'PAT' reading comprehension and vocabulary tests (NZCER, 1969); sentence-combining; error correction tests; literature appreciation tests and anonymous questionnaires to assess attitudes to work.

In the Myhill study a pre-test was administered to the pupils, and at the end of the study a post-test was given. The test was a piece of first person narrative "written under controlled conditions", encouraging the pupils to draw on their personal experiences. The test design and marking "were led by Cambridge Assessment". Each test was marked by two people, and a third marker resolved any differences. The markers did not know from which pupil group the pieces had originated (blinded assessment of outcome). The marking was based on the mark scheme format used by secondary schools at the time. The outcome was the change in the test scores using an ordinary least squares regression approach with pupil level data. One control class did not adhere to the intervention and was removed from the analysis.

Results and conclusions

The Elley intervention transformational grammar (TG) and 'Let's Learn English' (LLE) grammar groups found English more 'repetitive' and 'useless' than in the control group. The reading/writing (RW) group showed more positive attitudes to reading. The TG group were particularly negative about 'sentence-study'. In the fourth year (14/15 year-olds), only one comparison (from 30 possible) showed significant differences (on essay content). In the School Certificate Examination there were no significant differences between the three programmes. In the fifth year (15/16 year-olds) only two of the 12

variables listed showed any significant differences (sentence-combining test and English usage test). Again, in the School Certificate Examination, there were no significant differences between the three groups. Overall, Transformational Grammar and Traditional Grammar teaching showed no measurable benefits. Participants in the RW group, who studied no formal grammar for three years, demonstrated competence in writing and related language skills fully equal to that shown by the two grammar groups. Elley et al concluded that "English grammar, whether traditional or transformational, has virtually no influence on the language growth of typical secondary school students" (p. 18). Elley et al dismissed the idea of the introduction of grammar at primary level mainly based on developmental theory: "it seems most unlikely that such training would be readily applied by children in their own writing. Furthermore, the researchers' empirical findings do not support the early introduction of grammar" (p.18).

In addition to a wide range of findings that included analysis of teacher subject knowledge, the Myhill study found a "highly significant" positive difference in marks in favour of the intervention groups, and concluded that "this represents the first robust statistical evidence for a beneficial impact of the teaching of grammar in students' writing attainment" (p. 151). The authors also concluded that:

> "the study represents the first large-scale study in any country of the benefits or otherwise of teaching grammar within a purposeful context in writing. It stands in contrast to previous studies which were either small-scale (Bateman and Zidonis 1966; Fogel and Ehri 2000) or which investigated whether discrete grammar instruction improved writing outcomes (Elley et al, 1975, 1979), and is the only study of its kind conducted in England" (p. 161).

As we demonstrated earlier, it was not strictly accurate to claim that the Elley study used "discrete grammar instruction" as its comparator. The issue of scale is also interesting. It is true that the numbers of students involved in the Myhill study was the largest to date, but what is important is not the scale *per se* but the quality, and power, of the design of any study. Scale is also implicated in the consideration of the results of just one study versus the combined results of many studies, an approach that is at the heart of systematic review and meta-analysis.

Quality of the methodology of the studies

It is important to note that both studies were undertaken in secondary schools, not in primary/elementary schools, therefore their findings could not reliably be generalised to defend any decisions made for primary/elementary education.

The Elley study, as reported, had a number of limitations. Its design was a quasi-experiment as it did not use random allocation to assign students to classes. In addition, the sample size of this cluster trial was small and underpowered. It was also limited by the fact that it was undertaken in only

one school. Other issues that undermine the validity of the study include not stating how the students were allocated into the groups and not stating whether the tests were administered and marked blind to allocation to minimise potential bias.

The Myhill study, as reported, also had a number of limitations. The authors did not use an *intention to treat* method of analysis. Removal of a non-compliant class from the analysis *potentially* biased the results. This is because that particular teacher and class were likely to be systematically different from those who remained in the study. Randomisation ensures differences are balanced between the two groups *at baseline*. Removal of a class from one group, *post-randomisation* re-introduces the potential for the selection bias that the randomisation had previously dealt with by ensuring classes were similar between the two groups *at randomisation*. The second limitation is the bias in the standard errors. As the authors acknowledge, their study was a cluster randomised trial and, although they mention the need to adjust for clustering, they argue that, because there was only one cluster per school this was not necessary. Consequently, they treated the sample as having several hundred *independent* observations rather than 32 (or 31 after removal of the non-compliant teacher) *clustered* observations. Other issues that potentially threaten the validity of the study include: not describing who did the randomisation and not stating whether this was done independently of the investigators (developers); and not stating whether the pre-tests were done before random allocation to minimise potential bias from the participants having knowledge of the allocation before undertaking the pre-test. However, all of the limitations observed in the Myhill study were also possibly present in the Elley study but due to some limitations in the reporting of that study it is not possible to make a judgement about, for example, whether or not ITT analysis was used.

To conclude our in-depth analysis of the methodology of these single trials we look finally at a more recent study addressing the question of whether the Myhill et al contextualised approach was effective and generalizable to the oldest children in primary schools. The study by Torgerson et al (2014) was carried out as an independent follow-up trial funded by the EEF. This trial was aimed at pupils in the ‘transition period’ between primary and secondary school (last term of year 6 (age 10-11) and first term of year 7 (age 11-12)). The Torgerson et al study has not yet demonstrated similar levels of impact and significance as the studies by Elley et al and Myhill et al. but the comparison is relevant because the nature of the grammar intervention evaluated in this RCT was the Myhill approach. The inclusion of primary age pupils in the Torgerson study is also important for our argument in this paper, although the comparison with the Myhill study needs to be treated with caution due to the difference in participant characteristics.

The design of Torgerson et al study was a pragmatic ‘partial split plot’ randomised controlled trial. Schools were randomised at the cluster level (similar to the original Myhill design). In the intervention schools, children were additionally randomised as individuals to receive the grammar teaching as a whole class or in small groups. This allowed the evaluators to test whether

small group teaching was effective, as well as whether grammar teaching *per se* was effective. Unlike in the Myhill study the evaluators took the clustered nature of the data into account in the analysis and also undertook an intention to treat analysis, whereby all schools and pupils were included in the analysis, irrespective of their level of intervention compliance. The results showed that there was a small, statistically *non-significant* effect of grammar teaching on literacy outcomes. In contrast, the small group teaching delivered a modest, statistically significant effect on literacy outcomes. Indeed, when the small group effect was removed from the grammar teaching by comparing the whole classes in the intervention against the whole class control group the small difference declined from 0.10 of a standard deviation to 0.06. Therefore, the results of this trial suggest, at best, only a very small effect of grammar teaching on literacy outcomes. However, although the study did use an intention to treat analytical strategy, and the correct statistical approach, this was implemented among children during the 'transition' from primary to secondary school, which could have led to an underestimation of the teaching effectiveness, due to the summer break from attendance at school.

Discussion and conclusions

The last 30 years has shown a gradual increase in the use of experimental trials in education research. Greater understanding of the strengths and weaknesses of research designs is evident in more recent research studies. This greater understanding is reflected, for example, in the combining of experimental trials with qualitative methods including implementation process evaluations or embedded ethnography. In general these developments reflect growing sophistication in education research and social-science research more generally.

Although the numbers of robust experimental trials relevant to effective teaching in schools have increased, our analysis of trials in relation to the teaching of writing suggests that there are still too many studies that are not of sufficient methodological quality. In particular too many studies are weak in relation to allocation of pupils to groups, and the measures for writing remain a challenge. Randomisation, to form two or more intervention and control groups, is essential to ensure that the groups are balanced in known and unknown factors that may affect writing outcomes. Randomisation could be by pupil, by class, by school year or by school. The higher the unit of allocation (e.g., school versus pupil), the lower the efficiency (in statistical terms) of the design. In other words, all things being equal, it is necessary to have more children in a design that randomises at a level *above* the pupil to see a given difference (if one exists) that would be statistically significant. The main weakness of randomisation at the level of the child is contamination or spill over effects and the logistics of allocating pupils in ways that are different from the normal ways that schools allocate pupils to classes, hence the use of group or cluster randomisation of schools. A 'business as usual' control group is often appropriate in a pragmatic trial; however, it is useful to consider additional interventions leading to three or more arms to the trial if there were other competing interventions that could potentially improve writing skills.

Writing outcome measures ideally need to include robust measures for improvements in writing composition even if the focus is, for example, development of grammar. There is a need to know that the holistic aspects of writing are being enhanced not just the key components. Such an outcome measure should be administered and marked by independent assessors for whom the allocation of teaching approaches to groups is not known (the markers are 'masked'). This prevents either conscious or unconscious marking bias of the outcomes.

When the data are all collected and collated it is important to analyse the data as if *all* the pupils had received the intervention to which they were allocated whether they did or did not indeed receive the intervention (adopting 'intention to treat' or 'intention to teach' analysis: Torgerson and Torgerson, 2008). If schools that comply weakly with the intervention are excluded from the analysis this introduces the potential for selection bias which the original randomisation minimised. There are statistical techniques for looking at the effect of low compliance but removing weakly or non-compliant classes or schools is not one of them.

With regard to our substantive case of grammar, the current evidence from randomised controlled trials does not support the widespread use of grammar teaching for improving writing among native English speaking children. Based on the experimental trial and meta-analysis evidence about writing teaching more generally (e.g., in tables 1 and 2), our hypotheses are that supporting primary/elementary pupils' grammar is most likely to require teachers intervening during the writing process, and interacting to discuss the use of grammar in relation to the overall purpose of the writing task and the purpose of the writing. The necessity to use technical terms with pupils such as subordinate clause or subjunctive remains a question open to research, but it is doubtful that attention to such terms is beneficial. It is probable that adopting every-day language to discuss improvements in the use of grammar in writing will be more beneficial. Small group and whole class teaching that includes a focus on the actual use of grammar in real examples of writing, including professionally produced pieces, realistic examples produced by teachers including 'think aloud' live drafting of text, and drafts of pupils' writing, may also be more effective.

When the decisions taken by, and for, schools and teachers about what approaches to adopt are informed by research, there are important choices to be made. Although grammar for writing has been a main focus of this paper, if the overall goal is to improve pupils' writing then a much wider set of research evidence about writing needs to be considered. Improvements in pupils' writing have to be achieved across many different dimensions. For example, robust evidence has shown that an approach with primary age pupils, that used strategy instruction (itself an approach backed by robust multiple trial evidence) combined with pupils' experience of offsite visits to places of education interest, had powerful affects. This work had its origins in the US but an evaluation using RCT design undertaken in England confirmed its transferability to a different national context, although the trial was relatively

small and the results need to be confirmed in a larger effectiveness trial (Torgerson and Torgerson, 2014). However, once again this work is but one study and one approach. The most recent meta-analyses of high quality research studies on writing suggest that, rather than emphasise grammar, the following practices could be selected as a priority for teaching writing in primary/elementary education: a) an increase in the amount of time that pupils have for writing; b) adoption of a process approach to writing; c) creation of a classroom environment that is appropriately supportive of pupils' attempts at learning to write better; d) development of pupils writing skills, strategies and knowledge, including ways of planning writing; e) a use of assessment for learning techniques; f) a use of computers as part of the process of writing; g) a use of writing meaningfully across different subject areas (Graham, Harris & Chambers, 2016). The robustness of the evidence underpinning these practices is built not on single studies but on multiple RCTs and experimental trials.

The mismatch between curriculum policy for the subject English and the research evidence base is particularly pronounced at primary/elementary level in England. The national curriculum in England and its associated national statutory tests include a heavy emphasis on formal grammar teaching, and to varying degrees the national curricula in other English-speaking countries also have an emphasis on formal grammar teaching. Sentence-combining remains the only approach to grammar for writing that *is* supported by robust research evidence from experimental trials, although there are no RCTs that have been undertaken in the UK. The use of sentence-combining as part of the process of writing would be a good area for new research.

In relation to the use of evidence to guide policy, a key risk is for policy makers and their advisors is to attend too closely to single studies, within a field of interest, that might support a preferred policy direction rather than take due account of multiple studies published over many years. The problems of attending to a single study have been seen in relation to the teaching of reading in the UK (Ellis & Moss, 2013; Wyse and Goswami, 2008), and this, in addition to ideological belief, appears to be a reason for the dramatic emphasis on grammar in England's primary national curriculum that was implemented from 2014 onwards, a trend that is counter to the research evidence overall, and one that risks having a negative impact on children's literacy learning and hence life chances. The outcomes of reviews of multiple studies, including systematic review and meta-analysis and high quality narrative reviews, are a much more reliable evidence-base for policy decisions than single studies. But this kind of evidence also requires mediation by experts who possess both substantive, methodological and practical knowledge and experience.

Although policy makers and politicians around the world have engaged with the importance of research evidence, for example in the prioritisation of evidence-based practices based on RCTs, there is a resulting need for policy to accurately reflect the outcomes of robust reviews of multiple sets of evidence. Such reviews may indicate that a policy should be in a direction that is contrary to a minister's ideology and personal beliefs. At other times there

may not be sufficient research evidence to warrant a particular policy decision in any direction: in these cases there is the option to further prioritise schools' autonomy and teachers' professional judgement. Better policies are likely to be made in future if policy decisions are informed by expert critical synthesis of multiple robust research studies, including systematic reviews and meta-analyses, relevant to the contexts of implementation. Finally, a necessary consequence of the kind of attention to research evidence that we advocate may mean that curriculum policy should change more slowly and more incrementally because accumulation of the multiple studies that are required to warrant decisions in important areas such as the teaching of writing takes many years.

References

Andrews R, Torgerson C, Beverton S, Locke T, Low G, Robinson A, Zhu D (2004a) The effect of grammar teaching (syntax) in English on 5 to 16 year olds' accuracy and quality in written composition. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Andrews R, Torgerson C, Beverton S, Freeman A, Locke T, Low G, Robinson A, Zhu D (2004b) The effect of grammar teaching (sentence combining) in English on 5 to 16 year olds' accuracy and quality in written composition. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Australian Curriculum Assessment and Reporting Authority (Producer). (2017, 27 January 2017). Australian Curriculum: English. Retrieved from http://www.australiancurriculum.edu.au/english/structure.

Bateman DR, Zidonis FJ (1966) *The Effect of A Study of Transformational Grammar on the Writing of Ninth and Tenth Graders*. Champagne, IL, USA: National Council of Teachers of English.

British Educational Research Association (BERA). (2012). Background to Michael Gove's response to the Report of the Expert Panel for the National Curriculum Review in England. Retrieved from https://www.bera.ac.uk/promoting-educational-research/issues/background-to-michael-goves-response-to-the-report-of-the-expert-panel-for-the-national-curriculum-review-in-england

Bew, P. (2011). *Independent Review of Key Stage 2 testing, assessment and accountability. Final Report*. London: Department of Education.

Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science and Medicine, 75*, 2299-2306.

Connolly, P. (2015, September). *THE TRIALS OF EVIDENCE-BASED PRACTICE IN EDUCATION*. Keynote address at the British Educational Research Association Annual Conference, Queen's University Belfast.

Crystal, D. (2004). *The Stories of English*. London: Penguin/Allen Lane.

Department for Education. (2010). *The Importance of Teaching: The Schools White Paper 2010*. Norwich: The Stationery Office.

Department for Education. (2013a). *The national curriculum in England: Framework document. December 2014.* London: Department for Education.

Department of Education. (2013b). *Reform of the national curriculum in England. Report of the consultation conducted February – April 2013*. London: Department of Education.

Department of education. (2013c). *Reforming the national curriculum in England. Summary report of the July to August 2013 consultation on the new programmes of study and attainment targets from September 2014*. London: Department of Education.

Donner, A. and Klar, N. (2000) Design and Analysis of Cluster Randomization Trials in Health Research, London: Arnold.

Economic and Social Research Council (Producer). (2014, 27 January 2017). Improving literacy with grammar methods. Retrieved from http://www.esrc.ac.uk/news-events-and-publications/impact-case-studies/improving-literacy-with-grammar-methods/

Education Arena (ND). Expert Interview with Debra Myhill. Retrieved from: http://www.educationarena.com/searchResults/index.asp?cx=006001425124917276529%3A7n-kwidpf_c&cof=FORID%3A11&ie=UTF-8&q=hot+topic+debbie+myhill&sa=GO&siteurl=www.educationarena.com%2FeducContact.asp&ref=www.google.co.uk%2F&ss=5817j2852731j23

Education Standards Research Team. (2012). *What is the research evidence on writing?* London: Department for Education.

Education Endowment Foundation (EEF) (2017). The EEF's Approach to Evaluation. Retrieved from https://educationendowmentfoundation.org.uk/our-work/the-eefs-approach-to-evaluation/

Elley, W. B., Barham, I. H., Lamb, H., & Wyllie, M. (1976). The Role of Grammar in a Secondary School English Curriculum. Research in the Teaching of English. *10*, 5-21.

Ellis, S., & Moss, G. (2013). Ethics, education policy and research: the phonics question reconsidered. *British Educational Research Journal, 40*(2), 241-260.

Fearn, L. & Farnan, N. (2007). When Is a Verb? Using Functional Grammar to Teach Writing. Journal of Basic Writing, Vol. 26, No. 1, 2007

Fogel H, Ehri LC (2000) Teaching elementary students who speak black English vernacular to write in standard English: effects of dialect transformation practice. *Contemporary Educational Psychology* **25:** 212-235.

Foresight Mental Capital and Wellbeing Project (2008). Final Project report. The Government Office for Science, London.

Goldacre, B. (2013). *Building Evidence into Education*. London: Department for Education.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. Journal of Educational Psychology, 104(4), 879-896. DOI: 10.1037/a0029185

Graham, S., & Harris, K. (2017). Evidence-Based Writing Practices: A Meta-Analysis Of Existing Meta-Analyses. In R. Redondo & K. Harris (Eds.),

Design Principles for Teaching Effective Writing. Leiden: The Netherlands.

Graham, S., Harris, K., & Chambers, A. (2016). Evidence-Based Practice and Writing Instruction: A Review of Reviews. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research (2nd Edition)*. New York: The Guilford Press.

Graham, S., Bruch, J., Fitzgerald, J., Friedrich, L., Furgeson, J., Greene, K., Kim, J., Lyskawa, J.,, & Olson, C. B., & Smither Wulsin, C. (2016). *Teaching secondary students to write effectively (NCEE 2017-4002).* Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Graham, S. & Perin, D. (2007a). A Meta-Analysis of Writing Instruction for Adolescent Students. Journal of Educational Psychology, Vol. 99, No. 3, 445-476.

Graham, S., & Perin, D. (2007b). Writing next: Effective strategies to improve writing of adolescents in middle and high schools – A report to Carnegie Corporation of New York. Washington, DC: Alliance for Excellent Education.

Graham, S. & Perin, D. (2007c) What We Know, What We Still Need to Know: Teaching Adolescents to Write, Scientific Studies of Reading, 11:4, 313-335, DOI: 10.1080/10888430701530664

Hargreaves, D. (Producer). (1996, April 2016). Teaching as a research-based profession: possibilities and prospects, the Teacher Training Agency Annual Lecture April 1996. Retrieved from https://eppi.ioe.ac.uk/cms/Portals/0/

Harris, R. J. (1962) An experimental inquiry into the functions and value of formal grammar in the teaching of English, with special reference to the teaching of correct written English to children aged twelve to fourteen. PhD thesis, University of London.

House of Commons Education Committee. (2017). *Primary assessment. Eleventh Report of Session 2016–17. Report, together with formal minutes relating to the report*. London: House of Commons.

Johnson, B., Onwuegbuzie, A., de Waal, C., Stefurak, T., & Hildebrand, D. (2017). Unpacking Pragmatism for Mixed Methods Research. In D. Wyse, N. Selwyn, E. Smith, & N. Selwyn (Eds.), *The BERA/SAGE Handbook of Educational Research*. London: SAGE.

Kaestle, C. (1993). The Awful Reputation of Education Research. *Educational Researcher, 22*(1), 26-31.

Mansell, W. (Producer). (2017, 9 May). Battle on the adverbials front: grammar advisers raise worries about Sats tests and teaching. Retrieved from https://www.theguardian.com/education/2017/may/09/fronted-adverbials-sats-grammar-test-primary

Moore, L., Graham, A., & Diamond, I. (2003). On the Feasibility of Conducting Randomised Trials in Education: case study of a sex education intervention. *British Educational Research Journal, 29*(5), 673-689.

Morrison, K. (2001). Randomised Controlled Trials for Evidence-based Education: Some Problems in Judging 'What Works'. *Evaluation & Research in Education, 15*(2), 69-83.

Myhill, D., Jones, S., Lines, H., & Watson, A. (2011). Re-thinking grammar: the impact of embedded grammar teaching on students' writing and students' metalinguistic understanding. *Research Papers in Education, 27*(2), 139-166.

Myhill, D., & Watson, A. (2014). The Role of Grammar in the Writing Curriculum: A Review. *Journal of Child Language Teaching and Therapy, 30*(1), 41-62.

National Governors Association Center for Best Practices Council of Chief State School Officers (Producer). (2010, 27 January 2017). Common Core State Standards (English Language Arts Standards). Retrieved from http://www.corestandards.org/ELA-Literacy/L/5/

New Zealand Ministry of Education (2007). The New Zealand Curriculum: for English-medium teaching and learning in years 1-13.

Office for National Statistics. (2015). *UK Labour Market, February 2015*. London: Office for National Statistics.

O'Hare, F. (1973) Sentence Combining: Improving Student Writing without Formal Grammar Instruction. No. 15 in a series of research reports sponsored by the NCTE Committee on Research (Urbana, National Council of Teachers of English).

Pawson, R., & Tilley, N. (1997). *Realistic evaluation.* London: SAGE.

Saddler, B., & Graham, S. (2005). The effects of peer-assisted sentence-combining instruction on the writing performance of more and less skilled young writers. Journal of Educational Psychology, 97(1), 43–54.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA.: Wadsworth.

Slavin, R. (2002). Evidence-Based Education Policies: Transforming Educational Practice and Research. *Educational Researcher, 3*(7), 15-21.

Standards and Testing Agency. (2015). *2016 national curriculum assessments. Interim teacher assessment frameworks at the end of key stage 2*. London: Standards and Testing Agency.

Standards and Testing Agency. (2016). *key stage 2 English grammar, punctuation and spelling Paper 1: questions.* London: Standards and Testing Agency.

The Government Office for Science. (2008). *Foresight Mental Capital and Wellbeing Project (2008). Final Project report.* London: The Government Office for Science.

Torgerson, D. & Torgerson, C. (2008). *Designing and running randomised trials in health,* education *and the social sciences*. Basingstoke, Palgrave Macmillan.

Torgerson DJ, Torgerson CJ, Mitchell N, Buckley H, Ainsworth H, Heaps C, Jefferson L. (2014). Grammar for writing: Evaluation report and executive summary. *Educational Endowment Foundation* https://educationendowmentfoundation.org.uk/public/files/Support/Campaigns/Evaluation_Reports/EEF_Project_Report_GrammarForWriting.pdf

UNESCO (2006). *EFA Global Monitoring Report 2006: Education for All.* Paris: France.

UNESCO Institute for Statistics (Producer). (2016, 11 November 2016). Literacy. Available from http://www.uis.unesco.org/Literacy/Pages/default.aspx

United Nations (2017). Sustainable Development Goals. Available from http://www.un.org/sustainabledevelopment/sustainable-development-goals/

Walters, J.E. (1931). Seniors as counsellors. *Journal of Higher Education*, Vol. 2 (8) p. 446-8.

Walters, J.E. (1932). Measuring effectiveness of personnel counselling. *Personnel Journal*, Vol. 11, p. 227-36.

Wilby, P. (2017, 1 August). David Laws: 'The quality of education policymaking is poor'. *theguardian*. Retrieved from https://www.theguardian.com/education/2017/aug/01/david-laws-education-policy-schools-minister-thinktank-epi?CMP=Share_iOSApp_Other

Wyse, D. (2001). Grammar. For Writing?: A Critical Review of Empirical Evidence. *British Journal of Educational Studies, 49 (4)*, 411-427.

Wyse, D., Baumfield, V., Egan, D., Gallagher, C., Hayward, L., Hulme, M., . . . Lingard, B. (2013). *Creating the Curriculum*. London: Routledge.

Wyse, D., & Goswami, U. (2008). Synthetic Phonics and the Teaching of Reading. *British Educational Research Journal, 34*(6), 691-710.

Wyse, D., Fentiman, A., Sugrue, C. and Moon, S. (2014). English Language Teaching and Whole School Professional Development in Tanzania. *International Journal of Educational Development.* Vol. 38, 59-68. DOI: 10.1016/j.ijedudev.2014.04.002