University of Southampton

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

University of Southampton

Faculty of Environmental and Life Sciences

School of Psychology

Teacher Expectations and Language in a Classroom Context

by

Rebecca Sharon Thompson

ORCID ID 0000-0002-4358-2616

Thesis for the degree of Doctor of Educational Psychology

June 2023

Abstract

The systematic review explores the influence of pupil gender on the judgements of teachers. It considers whether pupil gender is a factor in teachers' decisions and expectations of pupil academic outcomes.

The findings from these studies indicated that there were gender-based differences in expectations which were linked with SEND, personal qualities and characteristics, and different academic subjects. Teachers also used gender-based reasoning to explain the causes of underachievement. Studies also indicated observable differences in teacher behaviours in the classroom. This review suggests that one reason for differing outcomes in such studies is the contrasting methodologies between using assessment data and using fictional vignettes.

The mixed method empirical study explores the nature of, and differences between teacher language use in whole class contexts compared with small groups. Observation data showed that there was a significant interaction between type of language used and the context. Semi-structured interviews explored teachers' thinking and attitudes about their language use and five analytic themes were generated. These are discussed with direct reference to the Stages of Change model (Prochaska & Di Clemente, 1982).

Helping children learn requires a breadth of thinking that includes adjusting the environment (e.g., making use of small group contexts) adapting language for individual pupils, and the careful use of non-verbal communication. A checklist resource is developed and offered to support teacher reflections on their use of language and to consider the reorienting of strategies to further think about enabling learning and promoting positive environments.

Table of Contents

Table of Tables

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Rebecca Sharon Thompson

Title of thesis: Teacher Expectations and Language in a Classroom Context

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signature: Date: 07/11/2023

Acknowledgements

I would like to offer my heartfelt gratitude to my two wise and generous supervisors who so willingly gave their time and expertise throughout the whole process: Dr Tim Cooke and Dr Fiona Okai. Thank you also to Dr Cora Sargeant and Isobel Brentnall for your help in this research.

I would also like to thank my amazing friends and family who have listened, supported, and sustained me on this journey. You know who you are, and I treasure you all.

To my parents, Bob and Margaret, and my sister Vicky who have a offered a lifetime of unconditional love and encouragement, I am eternally grateful for your endless support. You are truly special.

Words cannot express my immense gratitude to my wonderful husband Paul who believed in me every step of the way, even when I did not believe in myself. None of this would have been possible without your constant patience, your incredible support, your selflessness and your overwhelming love. And to Dante and Jasper – my amazing boys. You are an unending source of encouragement, inspiration and blessing and I could not have accomplished this without you. I cherish all that we are as a family and always will.

Definitions and Abbreviations

BPS ..... British Psychological Society

DfE ..... Department for Education

EHCP ..... Education, Health and Care Plan

ENCA ..... Evaluation of National Curriculum Assessment

FSM ..... Free School Meals

IRF ..... Initiation-Response-Feedback

LA ..... Local Authority

M ..... Mean

MICRO ..... Mixed Interval Class Room Observation

NASUWT ..... National Association of Schoolmasters Union of Women Teachers

NFER ..... National Foundation for Educational Research

OECD ..... Organisation for Economic Co-operation and Development

OFSTED ..... Office for Standards in Education, Children's services and Skills

OPTIC ..... Observing Pupils and Teachers in Classroom observation schedule

PRISMA ..... Preferred Reporting Items for Systematic Reviews and Meta-Analyses

SEMH ..... Social, Emotional and Mental Health

SENCo ..... Special Educational Needs Co-ordinator

SEND ..... Special Educational Needs and Disability

SPRinG ..... Social Pedagogic Research into Group-work

STEM ..... Science, Technology, Engineering and Mathematics

TEP ..... Trainee Educational Psychologist

Teacher Expectations and Language in a Classroom Context

1 - Reflections

Personal interest

Having been a primary school teacher for many years, I am very familiar with the assessment and data analysis required to ensure all pupils are achieving well. This included interrogating data to ensure there was a focus on vulnerable pupils or different groups e.g., confirming that boys and girls were achieving at comparable levels across all subjects. It was also important to know each individual pupil thoroughly and how I was intervening when necessary. Data analysis was also an important part of my later role as a SENCo where I tracked the gender breakdown of pupils requiring SEN support and those on an EHCP. Clearly, I was an important part of a system trying to ensure a focus on individuals and to openly track if certain groups were not achieving as they should. But therein lies the question. How should certain groups achieve? Do boys have strengths in certain subjects whilst girls have strengths in others? It made me question what expectations we, as education professionals, hold for our children and from where these arise.

I also remember many professional development sessions or books which stated that excellent teachers have high expectations of their pupils and being rather puzzled. What teacher would not have high expectations for their class? This sparked an interest in the Pygmalion research (Rosenthal & Jacobson, 1968) and the idea of manipulating teacher expectations which could in turn, have an impact on pupil outcomes. The work of Wang et al. (2018) summarised 30 years of teacher expectation research and suggested that the impact of pupil gender on teacher expectations was an area requiring further research. This then became the focus of my systematic review.

Teacher expectations clearly have the potential to impact teacher behaviour in the classroom. My interest in teacher behaviour was narrowed to a focus on their use of language by my awareness of the work of Harrop & Swinson (2000) and Apter et al. (2010) who broke down language into different categories. I wondered if teachers were aware of these categories or the

frequency with which they used different types of language. Certainly, teachers are taught to be positive in the classroom, and I wondered whether teachers reflect any further than this. Whilst teacher expectations may have a role in shaping teacher behaviours and language (as discussed in the systematic review) there may be many differing reasons underlying their approach. I wanted to discover and explore these further.

Wider literature

Current studies in the field offer statistical approaches to exploring language in the classroom. Increasing praise within the classroom has been the subject of research since the 1960s (Floress et al., 2018). Studies have pursued suggestions about optimal praise-to-reprimand ratios (Caldarella et al., 2020, 2021; Cook et al., 2017), training teachers to use more approval than disapproval in the classroom to increase pupil on-task behaviour (Swinson & Harrop, 2010), and have explored different types of praise; Apter et al. (2010) considered teachers' verbal behaviours in the categories of positive or negative, categorised whether the comment was regarding pupil academic or social behaviours, or whether it was a routine direction. Floress, Briesch, et al. (2021) defined their categories as behaviour-specific praise, general praise, or a reprimand, and Harrop & Swinson (2000) used the categories of academic or social, approval or disapproval, individual or group, whether the comment included a description of the behaviour and finally whether there was a redirection or instruction to a required behaviour. One study (Sazak-Pinar & Guner-Yildiz, 2013) also included the categorisation of teachers' non-verbal behaviours alongside their verbal comments and utilised the categories of approval or disapproval, and academic or social behaviours.

However, gaps remain. Firstly, I identified a paucity of research which seeks to understand whether language changes in certain contexts (such as small groups and whole class). Secondly, studies did not seek to explore why teachers use the language that they do or when and how different types of language may be more frequent. Understanding these reasons could have a substantial impact in being able to modify teacher language should that be helpful. Therefore, I did not limit the empirical paper to teacher expectations of pupils (the focus of the systematic

literature review) for an explanation of language in the classroom but used a mixed methods approach to exploring what teacher language was present in the classroom and what the reasons for this might be, as expressed by teachers themselves. The empirical study also included an important step between observations and teacher explanations of their language; teachers were asked to recall the different amounts of language they had used. This allowed further insight into how teachers reflect on and view their language in the classroom and enabled data analysis into the consistency of their reflections with observable data.

Previous research has found differences in the verbal behaviours of teachers towards younger and older pupils. Apter et al. (2020) report differences in the results from two previous studies (Apter, 2016; Apter et al., 2010) in which secondary school teachers were less critical of academic work than primary teachers but offered less praise regarding behaviour. Similarly, Sulla et al. (2019) found lower rates of approval and higher rates of disapproval in secondary schools compared to primary schools. For this reason, and my previous experience and familiarity with primary schools, I chose to focus on primary aged children in both papers.

Ontology and epistemology

It is important to consider ontology and epistemology when conducting research to ensure the methodology is appropriate and the study is theoretically sound (Crotty, 1998; Fryer, 2022b).

Ontology refers to a position about reality. A realist view argues there are real entities, objects and processes which exist independently of an observer and that it is possible to find them. However, an irrealist position states that reality is simply what people perceive to be real.

When considering epistemology, Fryer (2022b) describes two opposing ends of the epistemological spectrum (objectivist and subjectivist) and recognises that there are positions along the continuum. A subjectivist view of how knowledge should be gathered and understood suggests there is no one single reality or external truth; all knowledge is theory dependent and constructed by individuals in their environment. In this way, meaning is dependent on the person

constructing their truth and so understanding different perspectives is crucial. Contrastingly, an objectivist theory of knowledge argues that there is one single reality or truth and that it can be observed, measured or uncovered. This makes knowledge generalisable as it is possible to discover and apply rules.

The position in this research is that of realism and subjectivism, more commonly known as critical realism. The critical realist perspective recognises that there are aspects of the world which are 'real' and universal truths are accessible from an objective perspective. However, critical realism also acknowledges that uncovering and exploring reality is fallible, and that individuals are different and may interpret their experiences differently. It is therefore impossible to fully comprehend this reality as our perceptions and experiences are subjective. According to Fletcher (2017), humans can only ever capture a small part of an extensive reality. McEvoy & Richards (2006) argue that the goal for critical realist researchers is not to establish generalisable rules or to focus entirely on lived experiences and construction through knowledge and discourse; the main goal is to foster a deeper level of understanding and insight, in order to offer explanations.

Within this study, the real or objective phenomena were the comments being made by teachers in the classroom setting. However, I, through the lens of critical realism, recognised that what are defined as positive comments and what are defined as negative comments, along with memories and perceptions of the experiences during the lesson, are subjective. The mixed method approach of this study asked both what was happening in the classroom (the 'real' or objective phenomena) and why teachers were using the language they did (perceptions, memories and interpretations). Observation as a method, supported by a clear categorisation of comments codebook (see Appendix F), enabled collection of objective and observable data. Interviewing of teachers enabled a deeper understanding of their perceptions, beliefs and memories. Together these were consistent with the underpinning critical realist framework.

This study utilised a convergent mixed methods design in which I gathered both qualitative and quantitative data at the same time. I gathered and statistically analysed

quantitative data to understand the reality of differing rates of teacher language in the classroom whilst the semi-structured interviews with each individual teacher enabled me to explore aspects of their behaviour (choices, reasoning, explanations) that were not answerable from observations and quantifiable actions (Creswell & Creswell, 2022; Roer-strier & Kurman, 2009). In this way, I was able to use qualitative and quantitative methods in harmony, exploring the statistical data alongside the teachers' explanations in order to generate a deeper level of understanding, and to offer explanations, in line with the critical realist philosophical position. The minimal descriptive data examination used immediately before the interviews (tallying of types of comment and calculating the proportion) was designed to probe teachers' thinking and facilitate an insightful interview and was not an indication of beliefs in a data hierarchy.

Ethical considerations

Macfarlane (2009) posits six character virtues that researchers should consider when reflecting on their ethical practice which he argues can "empower rather than restrict discussion of key issues" (Macfarlane, 2009, p. 4). Each virtue aligns with a different stage of the research process: framing, negotiating, generating, creating, disseminating and reflecting. He suggests that these offer an alternative and broader approach to considering ethics in research than traditional codes of practice.

The framing stage of research refers to the initial development of the research questions and decisions around the study as a whole. The virtue required at this stage is courage, particularly the strength to risk failure. This was pertinent in this research when I was told by previous doctoral students not to pursue a mixed method approach. I was warned it was too involved and double the work. However, as my research question and initial reading of the wider literature developed it was clear that mixed method research would be the most appropriate way forward. Therefore, I had to be courageous in selecting the methodology for my research and philosophical position.

When at the negotiating stage, Macfarlane (2009) suggests that respect is the main virtue. I aimed to be respectful by going over and above the required BPS standards (e.g., informed

consent, right to withdraw, confidentiality). I shared with participants that I held Qualified Teacher Status and would observe in accordance with NASUWT union guidelines. I was also concerned that teachers may feel obliged to participate because of the monetary incentive for schools. (I had offered £100 to the Parent Teacher Association of each participating school. Further details are available in the Method section.) I therefore reassured teachers that the school would not lose the money if they withdrew from the study. I was also very open with participants regarding confidentiality and anonymity. Whilst I could reassure them of my confidentiality, it was possible we would be seen through a window during an observation and other staff would be aware of their participation, and therefore they would not be anonymous. When considering different levels of power within the study, I felt it was important to reassure teachers that their headteachers would not receive individual data, only the final write up.

During the generating stage of the research, Macfarlane (2009) asserts that researchers must be resolute. He describes the intense nature of doctoral studies and the strength of character required to complete research, also recognising that mature students, although bringing both professional experience and life experience, have other commitments such as children. I did indeed have to demonstrate resoluteness and determination throughout the research process. One example of this was when two of my pre-arranged schools were too busy to participate and stopped responding to emails. Advice from a colleague suggested giving them a deadline and then moving on to ask a different school to participate which I did.

The virtue of sincerity is at the heart of Macfarlane's (2009) creating stage. I felt that member checking following the thematic analysis demonstrated respect and sincerity. Participants were pleased to be consulted and valued being asked to ensure their contribution was reflected in the final themes. This meant I was returning a true picture of the thoughts they had shared. I was also sincere and honest during the interview stages, sharing the data I had collected with accuracy. This was sometimes tricky if I knew the data showed teachers being a little less positive than they hoped. However, I was clear, professional and ensured they viewed the data as part of a picture, not a judgement. I also felt that using the comment guidance from

Merrett & Wheldall (1986) when gathering observational data offered a clear foundation on which to build, thus ensuring accuracy and reliability in data collections across the different classrooms. (See the subsection 'Materials' in the Method section, supported by Appendix F, for details of categories, and the limitations section of chapter 3). I was disappointed that my research assistant was unable to join me for classroom observations which meant it was not possible to consider compare decision making at the observation stage.

I have not yet (at the time of writing) been able to offer any dissemination of this study. However, I have a presentation at a conference planned, I have promised to share the study with all participating schools, and I hope to publish both research papers in the Social Psychology of Education journal. Throughout the research process and on into the dissemination stage, I have sought to demonstrate the virtue of humility, respecting the research and work on which this study builds. During the semi-structured interviews, I was very clear in checking my understanding with teachers and asking if I was correct, ensuring I was not guiding their answers.

Reflexivity (the virtue required for Macfarlane's (2009) final stage 'reflecting') has played an important role throughout the research process from initial discussions with experienced supervisors and adapting the design, to amending the language used in light of the teacher's comments in the pilot study. I also recognise that there may be noise in my data which may have come from mistakes or judgements I made when observing in the classroom. Similarly, the data may have told a slightly different story if I had chosen to include non-verbal communication. If I was to complete the study again, I would seek to include a second researcher to offer the opportunity for greater collaboration during observations. Alongside this, a larger pilot study would give a broader range of comments to consider how to code. Particularly the repetition involved in small group phonic teaching in some classes led to some very quick 'in the moment' decision making. Written definitions of types of language would also help to ensure consistency across teachers and interviews and therefore offer even greater confidence in any conclusions. I believe there is also a place for non-verbal and para-verbal data within these types of studies. Whilst I did not include them in observation data collection, they too contribute to the overall

ethos and environment within the classroom and clearly teachers are aware of the power of these additional tools when communicating. Finally, in my efforts to be unobtrusive and not an inconvenience to schools, I observed whichever lesson was happening on the morning I was in school. I believe there is scope here for variation in language due to the lesson rather than the context as I observed maths, English, drama, PSHE, free flow play, topic and phonics sessions across the schools. I also observed some transition times within the observation hour in some classes. Therefore, I would consider always asking for a certain lesson or a certain time of day if I was to do this again.

Myself as a researcher

The two most significant things I learnt about myself as a researcher were my theoretical philosophy and my ability to demonstrate research integrity despite recognising a clear bias.

I am very aware that I worked in schools for 18 years before becoming a TEP and I have many friends who are teachers or Headteachers. I recognised that I hold a very strong bias in my passion and admiration for those working in the education system. Noting down accurately when they used redirective or negative language or having to tell them if they had been less positive than they hoped was really difficult. Nevertheless, I recognised this bias and the importance of research integrity and recorded or shared with honesty. I also ensured that I was professional and sensitive at all times and offered debriefings and follow up information in line with my ethical procedure.

Structure

This thesis is broken down into three chapters. This first chapter offered reflections on the process of conducting and writing the thesis including ethical considerations and the unique contribution of this research within the wider literature. Chapter two is a systematic literature review arising directly from the recommendations of a 2018 systematic review paper about teacher expectations in the classroom. Chapter three details an empirical study exploring teacher

language in the classroom including content and frequency in different contexts, predictions, and a thematic analysis which considers teachers' own explanations around the data.

Chapters two and three were not researched and written sequentially due to the constraints of the doctorate course and the need to work on both chapters in parallel. This meant that the systematic literature review was not completed before work began on the empirical paper. Initial reading for the systematic review identified gender as a potential influencing factor and this was used to refine the focus and scope of the review. The empirical paper retained its wider focus on teacher expectations and the language used in classrooms, without specific consideration of whether gender had a role. This is raised as a limitation in the strengths and limitations section of chapter three.

2 - A Systematic Literature Review Exploring the Influence of Pupil Gender on Teacher Judgements and Expectations of Pupil Outcomes

Abstract

This report details a systematic review which explores the influence of pupil gender on the judgements of teachers. It considers whether pupil gender is a factor in teachers' decisions and expectations of pupil academic outcomes.

The researcher used a systematic search strategy utilising seven databases in August 2022 and September 2023 to identify seven UK based studies which investigated this question. Inclusion criteria cited UK based, peer reviewed studies of the primary age range. Exclusion criteria included studies of a medical nature, pupils with specific diagnoses, and papers which considered pupil outcomes without consideration of initial teacher expectations. Of the seven studies eligible, five were exclusively quantitative, one was qualitative, and one was mixed methods. Participant numbers within studies ranged from 15 to 9610. This systematic review employed narrative synthesis to explore and present the heterogeneous studies.

The findings from these studies indicated that there were gender-based differences in expectations which were linked with SEND, personal qualities and characteristics, and different academic subjects. Teachers also used gender-based reasoning to explain the causes of underachievement. Studies also indicated observable differences in teacher behaviours in the classroom.

This review suggests that one reason for differing outcomes in such studies is the contrasting methodologies between using assessment data and using fictional vignettes. A limitation of the studies identified through quality assurance processes included a risk of bias within samples.

Further research into intersectionality effects would be beneficial along with wider studies to examine between-teacher or between-schools effects and to explore whether teacher

expectations adapt over time. Implications for practice including initial teacher training, national resources and foci, and the importance of individualised approaches are discussed.

Historical Context

The seminal study 'Pygmalion in the Classroom' (Rosenthal & Jacobson, 1968) is now over 50 years old yet the study of teacher expectations remains a thriving and continually researched area of psychology in which there is much still to learn. Teachers' expectations can be defined as "inferences that teachers make about the future behavior [sic] or academic achievement of their students, based on what they know about these students now" (Good and Brophy, 1997, p. 74). The initial 'Pygmalion' research suggested that it was possible to increase demonstrable outcomes for pupils through increasing teacher beliefs that high levels of success were imminent for particular children. The authors found that teacher expectations could have an impact on the outcomes of pupils; the beliefs of teachers (even if false) had the potential to create a self-fulfilling prophecy. In the subsequent plethora of expectations research, there is further evidence that teacher expectations positively predict pupil outcomes (Alvidrez & Weinstein, 1999; Rubie-Davies et al., 2014; Tandler & Dalbert, 2020; Wang et al., 2021).

Jussim (1986) proposed that there were three major stages to consider when studying self-fulfilling prophecies within a classroom context. Firstly, that teachers form inaccurate expectations, whether too low or too high. Secondly, that teachers treat pupils differently based on these expectations, and thirdly, that pupils respond in a way that confirms the initial expectation. The first of these stages is supported by meta-analysis of 75 studies across a variety of school types, pupil ages and subjects, in which the authors found that, whilst correlation between teacher judgements and pupils' test performance was "fairly high (.63)" (Südkamp et al., 2012, p. 755), there was space for the accuracy of teachers' judgements to improve. Similarly, recent research in Germany, studying 1065 first-grade pupils, found that "teacher expectations differ from actual student achievement" (Gentrup et al., 2020, p.12) and are partly inaccurate. They state that this is in line with existing research. However, Jussim himself, in a later study, offers a caveat, suggesting that it is possible for teachers' attainment predictions to be confirmed, not because the expectation was inaccurate and led to a self-fulfilling prophecy, but because the

teachers were able to skilfully predict the level of achievement for each pupil. They demonstrated "accuracy of prediction without influence" (Jussim, 1991, p. 67).

The second suggested stage in a self-fulfilling prophecy is that teachers treat pupils differently based on inaccurate expectations. Suggestions for some of the ways in which this might be happening include thousands of non-verbal communication cues (e.g., posture, facial expressions, raising of eyebrows) from teachers which pupils interpret and then behave accordingly (Zajda, 2021). When considering classroom specific practices, Good (2014) argues that teachers offer more praise, more rephrasing of questions, more opportunities and time to respond to questions, and more follow-up questions to pupils for whom they hold high expectations. Where lower expectations are held, teachers give more criticism, fewer opportunities and choice, and are quicker to provide an answer or move on to another pupil. Despite these findings, Gentrup et al. (2020) argue that there is relatively little evidence for this second stage, and it is still not clear exactly how teachers are communicating their expectations.

Jussim's (1986), third stage states that pupils respond in a way which confirms the initial expectation. However, Trouilloud et al. (2002) argue that teachers' final evaluations of pupils may be biased, not because pupil attainment has been influenced by those expectations, but because the original bias is still present. Similarly, confirmation bias may be evident throughout the process, in which teachers are more alert for pupil behaviour which matches their expectation and less likely to notice or remember behaviour that contradicts their initial judgement (Good & Brophy, 1994).

Jussim & Harber (2005, p. 131) argue that teacher expectation effects are often small and "that they may predict student outcomes more because these expectations are accurate than because they are self-fulfilling." Indeed, there are still unanswered questions.

Sources of Influence in the Formation of Teacher Expectations

As research moved beyond individual self-fulfilling prophecies, the focus shifted to whether there were group level characteristics which influenced teacher expectations. Many

areas have been suggested and researched as possible sources of influence in the formation of teachers' expectations of pupils. In a systematic review of teacher expectation literature (Wang et al., 2018), four analytical themes were identified across 142 studies. Theme one considered "influential factors on the formation of teacher expectations" (Wang, Rubie-Davies and Meissel, 2018, p. 129) and included research into the influence of areas such as ethnicity, gender, socio-economic status, immigration background, learning disabilities, emotional or behavioural difficulties, speech sound disorders, motivation, classroom engagement, and pupil work habits, on teachers' expectations. The other three themes considered in the systematic review were: mediating methods of teacher expectations; moderators of teacher expectation effects, and teacher expectation effects on student outcomes.

Gender Inequalities

Timmermans, Rubie-Davies and Rjosk (2018) highlight that evidence around inequalities in teacher expectations in some of the potential influencing areas (e.g., SES, SEND) is recognised and accepted, however evidence regarding expectations based on gender and ethnicity remains variable and inconsistent. Indeed Jones and Myhill (2010) claim that beliefs regarding gender have been used historically to account for under-achievement in males and females yet have also been used to explain achievement for both males and females. This is in line with the conclusions of the 2018 systematic review.

Statistical data demonstrate that there are gender differences in assessment scores across countries within the Organisation for Economic Co-operation and Development. When starting school, similar attainment is observed across genders. However, by the fourth grade (age 9) girls score higher than boys in reading, whereas boys outperform girls in maths. These gender differences in scores increase as pupils get older (Organisation for Economic Co-operation and Development, 2016). It may be that gender specific expectations of teachers play a role in stimulating, sustaining or exacerbating gender achievement gaps. Indeed some studies have found that teachers overestimate the language abilities of girls (Hinnant et al., 2009), boys are overestimated in mathematical ability (Holder & Kessels, 2017; Tiedemann, 2000), girls perceive

there are lower expectations on them for maths (Lazarides & Watt, 2015) and boys perceive there are lower expectations on them generally (Åhslund & Boström, 2018). But in further evidence of the complexity of this research area, other studies have found differing results; Gentrup & Rjosk (2018) found no significant interaction between teacher expectations and pupil gender when considering reading or maths in six-year-olds, whilst Kaiser et al. (2017) suggest that minority characteristics (of which one was gender) can actually increase teachers' accuracy of expectations.

Therefore, when considering future directions, the authors of the systematic review (Wang et al., 2018) suggest that research could potentially focus on a theme or sub theme in order to closely consider contextual factors, methodologies and the controlling of variables with the aim of clarifying some of the contradictory research findings. This research considers these suggestions and builds on the systematic review. It lies mostly within the first stage of Jussim's (1986) three-stage process hypothesis, exploring the formation of inaccurate expectations and specifically whether gender has a role or influence when teachers are forming judgements or considering potential outcomes for pupils. Some studies within this review offer evidence of stage two of the hypothesis, whereby there is consideration of teachers' observable behaviours.

Aims of Review and Why it is Important

The objective of this review is to explore the influence of pupil gender on the judgements of teachers. More specifically, it asks whether pupil gender is a factor in teachers' decisions and expectations of pupil academic outcomes and general classroom behaviour. As teacher expectation research has grown, understanding of the many and varied sources of influence, transmission, impact and effects has become both broader, confusing and with contradictory conclusions within certain research areas (Wang, Rubie-Davies and Meissel, 2018). Gendered expectations is one of these areas.

Self-concept refers to a set of beliefs held about the self and includes the ideal self, current self-image and self-esteem (Rogers, 1959). These self-evaluations are developed through internal and external comparisons and through gauging the reactions of significant people

including teachers (Harter, 2006). Sinclair et al. (2005) argue that people's self-concepts are malleable and will shift depending on the opinion of oneself held by others. This shift can be positive or negative but negative expectations have a stronger effect than positive expectations on the performance of high school students (Babad et al., 1982) and the link between early expectations (aged four) and later assessment data (aged 18) is strongest for those who have been underestimated (Alvidrez & Weinstein, 1999). Therefore, it is important to understand more about all aspects of teacher expectation nuance in order to positively intervene and support the best possible outcomes for pupils.

This systematic review employed narrative synthesis to explore and present the heterogeneous studies in a coherent and comprehensive account. An inductive approach was utilised with the intention of considering whether there were patterns or themes within the studies. This bottom-up approach meant that the strands and headings used in the synthesis of findings (SEND, personal qualities and characteristics, cause of underachievement, achievement in different subjects, observable behaviours) were derived from the researcher's reading and interpretation of the seven studies.

Method

Search Strategy

Initially, the search terms were taken directly from (Wang et al., 2018) in order to follow up their recommendation for further research. However, initial scoping searches showed that restricting searching to 'students' removed the possibility of finding papers which used synonyms e.g., pupils. Therefore, the word student was removed from the search and not replaced, so all possible synonyms were found. As this review was to explore the impact of gender specifically, search terms referring to this area were included using AND. In order to check the sensitivity of the search, the results were checked against those in theme one (influential factors on teacher expectations) of the review by Wang, Rubie-Davies and Meissel (2018). Seven databases were searched on 17th August 2022 and on 25th September 2023: PsycInfo, Medline, PsycArticles,

CINAHL, Web of Science, ERIC and Pro-Quest. These are commonly used databases in educational psychology.

As there were a high number of results, further filters were applied. Firstly, to restrict results to English language only and secondly to restrict results to primary aged children. In five databases this was achieved through filters (grades 1-7, elementary, and primary), and in two elementary, and primary were added to the search using AND.

This research then utilised the inclusion criteria of UK studies only. This was in line with the recommendation of Timmermans, Rubie-Davies and Rjosk (2018) who suggest that future research in this area needs to be ecologically sound, consider the context in which children and teachers are interacting, and the characteristics of classrooms and schools. For full search terms, see Appendix A.

Inclusion and Exclusion Criteria

1387 articles were identified through the search strategy and collated. Next any duplicates were removed, and an initial examination of titles and abstracts was conducted using predefined inclusion and exclusion criteria (see Table 1). Where abstracts did not specify the location of the study, the methodology section was read to ensure correct application of the inclusion and exclusion criteria. Full texts of the remaining articles (N = 9) were retrieved and read in full. One article did not meet the requirements of the review because it considered the impact of the teacher's judgement on the pupil's self-confidence rather than whether gender had impacted this initial judgement. The other excluded article was an opinion-based piece. Seven articles were taken forward for quality analysis and subsequent inclusion in the systematic literature review. The full process can be found in the PRISMA flow diagram (Figure 1). The review itself was not registered with a database.

Table 1

Inclusion and Exclusion Criteria

Include	Exclude
Papers directly investigating the influence or moderation effect of gender on how teachers make judgements or predict outcomes	1. Papers investigating the outcomes of teacher decisions but not what may have influenced decisions or judgements initially
English language	2. Not English language
	3. Studies in a medical field e.g., obesity, nutrition or smoking
	4. Studies into specific populations of pupils e.g., those with a diagnosis of autism, ADHD, language disorder, intellectual disability or other Special Educational Need and Disability (SEND).
	5. Studies which do not predict academic outcomes (e.g., bullying, homelessness)
UK based papers	6. Studies from outside of the UK
Studies focused on primary aged pupils	7. Studies with pupils under 6 or over 12 as the main population
Peer reviewed empirical research study either published in an academic Journal or having	8. Opinion or editorial based pieces

been peer reviewed through an academic

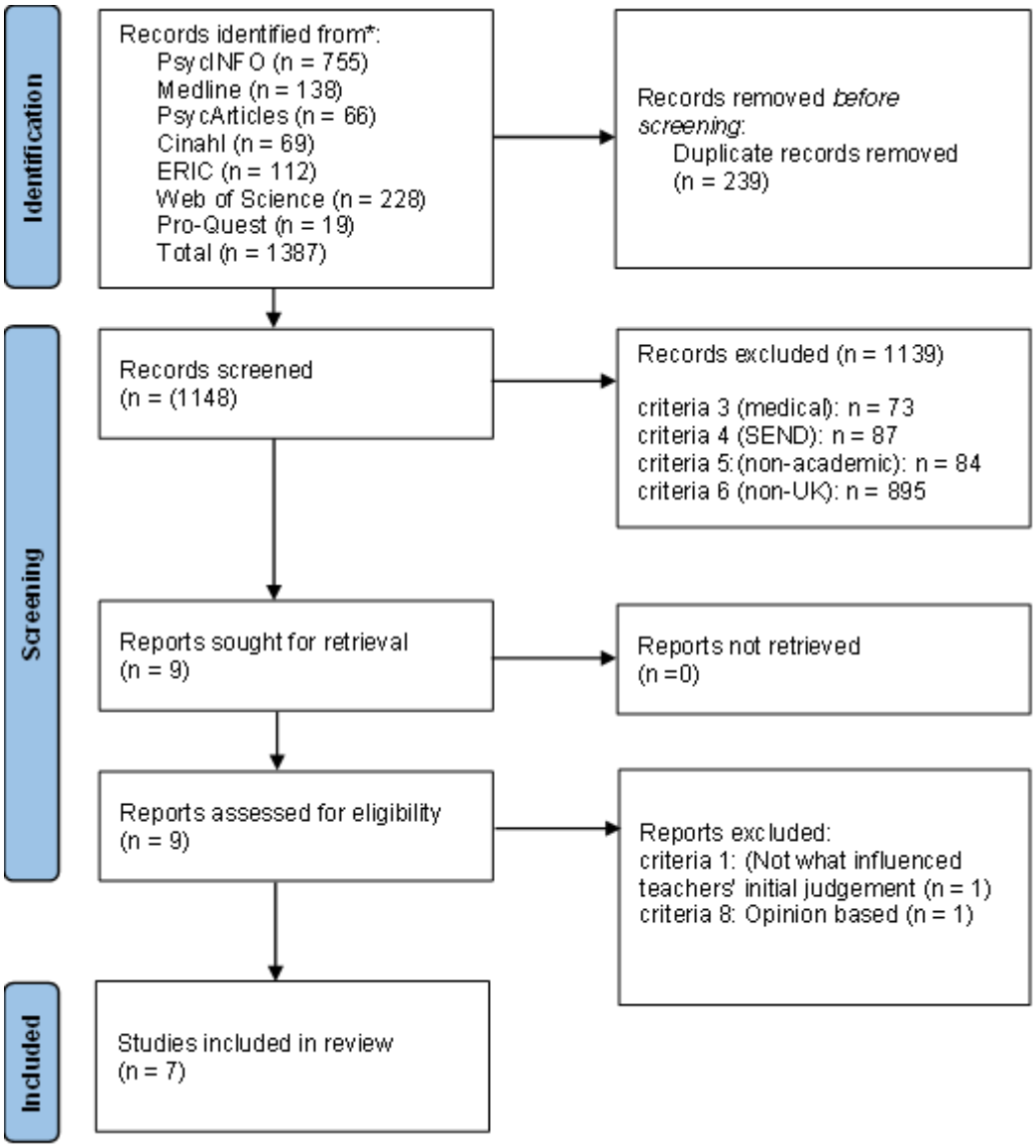
route (e.g., theses)

Quality Assessment

Across the seven studies there was a mix of quantitative, qualitative and mixed methods research. Two different quality assurance checklists were used with the mixed methods study being assessed on both. Quantitative studies were evaluated using the Critical Appraisal Framework (Woods, 2020) originally used in Bond et al. (2013) and later developed and fully explained in Flitcroft and Woods (2018). The checklist was initially designed as a scoring system but in line with Boland, Cherry and Dickson (2017), numerical scoring was changed to a tick system in order to support reflections on individual elements, some of which may be more important than others and require greater weight when evaluating the quality of the research. This meant that there was no numerical threshold at which studies were considered high, medium or low quality. However, the quality assurance checklist was used to gain a sense of the broad strength of the field and to guide the critique of individual papers. Neither checklist was given precedence over the other. Qualitative studies were evaluated using the Critical Appraisal Skills Programme checklist (CASP, 2019) which has been found to be a good measure of reporting standards and transparency (Long et al., 2020). One researcher quality assessed all seven studies and details can be found in Appendices B, and C. Both quality assurance checklists were broad enough in their questions for the range of methods across the seven studies whilst still enabling useful comparisons.

Figure 1

PRISMA Flow Diagram



Results

Study Characteristics

Seven studies met the criteria for inclusion based on the search strategy and criteria described. There was no date restriction within the search strategy and the publication dates of the studies spanned thirty years ranging from 1988 to 2018. Five were exclusively quantitative, one was qualitative, and one was mixed methods. The number of participants in each study varied widely from the smallest number at 15 (13 adults and two pupils) up to 9610. One researcher

read and collected data from each report. A data extraction table with detailed study characteristics can be found in Appendix D

Whilst each study aimed to explore the influence of gender on teacher expectations of pupil outcomes, three studies investigated this through statistical comparisons of standardised data with teacher assessment data, one used vignettes to probe teacher responses to fictional pupils, two utilised semi structured interviews of teachers, and one considered how teachers decided upon the ability group or 'set' for individual pupils and what factors may influence this decision. This was achieved through the use of standardised measures for different potential predictor variables such as behaviour, attitude and gender.

The narrative synthesis of findings of these heterogeneous studies is presented in line with Jussim's (1986) three stage model. This model provides a clear framework for discussion through a simple separation of the different potential stages in the process of self-fulfilling prophecies. Whilst Kelley (1992) argues that the separation of phenomena into smaller categories risks losing the overall perspective offered by molar analysis, this broken-down, linear approach was considered helpful for synthesising the variety of quantitative and qualitative data in the included studies. Discussion of the first stage includes possible influencing factors (SEND, personal characteristics and qualities, beliefs around typical and atypical behaviours, and beliefs regarding different subject strengths) which may contribute to the formation of inaccurate expectations. Discussion of the second stage concerns what evidence of these expectations there may be in teachers' observable behaviours and how teachers explain the behaviours and choices they make in the classroom. The third stage, in which pupils respond accordingly, is not the focus of this research although pupil outcome measures are frequently used in the literature as the measure by which to conclude if teacher expectations and judgements were accurate. The quality, methodology and strength of the research is then discussed.

Synthesis of Findings

Stage One: Teachers Form Inaccurate Expectations. In Jussim's (1986) first stage, the formation of inaccurate expectations occurs due to prior knowledge of the pupil before meeting

them, minimal or shallow knowledge of the pupil's characteristics or small amounts of initial assessment information. Studies in this review found inaccurate expectation development in four areas.

Gender and SEND. A study by Cline and Ertubey (1997) considered whether gender influenced teacher perceptions when considering children's difficulties. 529 teachers from 79 primary schools in the Home Counties responded to a questionnaire around support needs for different children. Vignettes, described as rounded and providing a fuller context than in previous studies, were given to teachers, who had to rate the level of support, intervention or graduated response which would be required to help the child. The options ranged from informal discussion with colleagues (universal or quality first teaching) to a formal Statement (as this level of support was known before EHCPs) and education in another school. The vignettes were identical for all teachers with the variable of gender changed in 50%. When analysing scores, it was the greatest three levels of graduated response which had high reliability scores and were analysed further. However, they found that judgements on the need for high levels of support and external expertise were not influenced by the gender of the child. This appears to contrast with the findings in Hamilton & Jones (2016) who interviewed two primary school teachers in north Wales. Despite this small sample size (due to it being a pilot study) they reported that one teacher often explained difficulties in the classroom for boys through the use of SEN syndromes, labels or disorders in a way that did not happen for female pupils. The researchers argue that despite this being a potentially positive and supportive response (indeed Ho (2004) argues it can help children, parents and teachers to understand and cope), it can unnecessarily pathologise children and impact their education in a number of different ways. These include feelings of shame and internal conflict (Kildea et al., 2011) increasing division, and promoting a focus on failures and impairments (Demetriou, 2020). Jones and Myhill (2010), who conducted research across 10 self-selecting schools and 40 teachers in the southwest of England, found more than one respondent who felt boys were more likely to have SEND. However this figure is difficult to consider further as the quality assurance process highlighted that there is no clear statement of how the interviews

were analysed by the researchers. An important difference between the studies was that Cline and Ertubey (1997) were asking teachers to consider fictional pupils whereas Jones and Myhill (2010) and Hamilton and Jones (2016) focused on reflections of real experiences.

Campbell (2015) used the Millennium Cohort Study (wave four) which offered a large sample size (girls = 2494 boys = 2491) and compared teacher judgements of pupils' ability with scores from standardised tests to consider potential biases in expectations. Here, there was a stronger bias in judgements for boys with SEND. Despite boys overall being more likely to be judged 'above average' in maths, a boy with SEND was judged more harshly than a girl with SEND when focusing on mathematical ability. These studies all suggest that gender does have an impact on the formation of teacher expectations, particularly when intertwined with SEND. However, it is also important to consider whether publicity and social construction around boys' learning needs has led to greater numbers of premature or unnecessary labels and impacted teacher expectations.

Gender and Personal Qualities and Characteristics. Doherty and Hier (1988) followed up an earlier investigation (Doherty & Conolly, 1985) in which pupil sex and tidiness interacted and played a role in teacher expectations, in that estimations of academic scores for tidy boys were higher than those of untidy boys. No similar differences were found for girls. In this later study (female teachers = 5, boys = 64, girls = 57), a personal rating scale (which included bi-polar constructs such as industrious or lazy, mature or immature, placid or anxious, reliable or unreliable, tidy or untidy) was used to measure teacher perceptions of pupils. Results showed that character and personal qualities became intertwined with academic ability and that this was more significant for boys. Negatively perceived children in relation to the personal rating scale became negatively perceived in terms of academic ability. This entanglement was especially severe for boys. However, those pupils who were rated as more tidy, more mature, more reliable, more placid, and/or more industrious were more likely to be predicted higher academic scores, regardless of ability. Doherty and Hier (1988), when reflecting on their results, ask what it is that makes boys so open to disadvantage when teachers are predicting their academic performance?

This raises concerns that teacher tolerance of aspects of personality such as tidiness are influencing their expectations about academic ability with boys being subject to harsher judgements.

Hartas (2018), used the Millennium Cohort Study of pupils (n=9610, boys = 50.9%) to investigate the likelihood of being placed in the middle or lower set at 11 years old. The research summarises a number of factors which appear to be strong indicators of a pupils' set position. Alongside gender, other teacher perceptions (e.g., of behaviour and attitude) and pupil characteristics (e.g., family income) were strong predictors of set position.

In the study by Campbell (2015) which looked at reading judgements by teachers of their pupils (girls = 2503 boys = 2494), there was an interaction effect between low income-level and gender, and membership of an ethnic minority and gender, which appeared to be stronger for girls. Girls were less likely than boys in these categories to be rated 'above average' for reading.

Gender and Underachievement. One study (Jones & Myhill, 2010) explored how teachers explained causes of underachievement in pupils. In the study, underachievement was defined as pupils who were "not achieving in academic tests" but that teachers believed had the ability to do better because they demonstrated other qualities such as "good general knowledge" or an ability to "grasp ideas and principles quickly" (Jones & Myhill, 2010, p. 552). The authors found that teachers had different explanations for boys' and girls' underachievement. These teachers articulated beliefs in which boys are perceived in a negative light of things they cannot or will not do and this contrasts with their opinions of girls; they do achieve. 80% of teachers in this study (n=32) believed that boys and girls should achieve the same results and they suggested that children have equal academic potential regardless of gender. However, when asked about typicality, the dominant responses were that a high achieving girl is typical and an under achieving boy is typical. In fact, high achieving boys were considered 'atypical,' with one teacher suggesting that a male pupil was atypical because he was well-mannered, articulate and had a work ethic and another suggesting that a particular boy was typical because he was silly, not because he was bright. Teachers also put forward that boys are active, not passive and they will challenge rather

than accept, and that this contributes to their underachievement. However, the researchers note that these could easily be given as reasons why boys do well and are not necessarily certainties for either underachieving or highly achieving. Similarly, the descriptions of outgoing and needing to be challenged could be argued either way. However, boys were also described as disruptive and not being able to sit still, which may have a stronger link to underachievement.

Causes of underachievement in girls were not widely discussed by teachers in the Jones and Myhill (2010) study and pupils themselves were more likely to comment on this subgroup than teachers. However, the main reason given was that girls may lack confidence, for example being less inclined to take risks. Poor behaviour in girls or a disaffected attitude was not specifically articulated by teachers despite the researchers themselves interviewing a girl who told them she did not want to be clever as this meant she'd have to do harder work.

Only a small minority of teachers rejected the idea of 'typical' in terms of gender but overall, Jones and Myhill (2010, p. 560) conclude that teachers have a "strong set of concepts and opinions" regarding underachieving boys. In summary, this paper found evidence of differing expectations for girls and boys in discussions around behaviour, attitude, and typical behaviours but was limited by its lack of reflection on its own methods, the selected methods (e.g., frequency tallies for interview comments), and a very broad research aim, all of which were highlighted by the quality assurance process, and mean that the findings of this paper should be treated with some caution.

Gender and Achievement in Different Subjects. Six of the papers considered how judgements for girls and boys varied across academic subjects. In a study focused on Standard Attainment Test (SAT) results at Key Stage one, Plewis (1997), compared teacher assessment (TA) levels with SAT results for 7400 pupils across 330 schools nationally. As expected, the chance of getting a high TA level increased as SAT score rose. However, when comparing across genders, the chance of girls receiving a higher TA level than their SAT result was 20% higher than for a boy. This pattern was seen across English, maths and science. Conversely, the other five studies which investigated this area did not find similar results spanning the different subject areas. Instead,

they reflect on differences in expectations for maths and English depending on pupil gender. Although these other five studies use a variety of methodologies, the overarching finding is that boys are expected to do better in maths and girls in English.

To consider maths achievement, Campbell (2015) compared teacher judgements of pupils' ability (n= 4985) with scores from a mathematics test. Predictor variables for whether a pupil was judged above average in maths were their score on the test and their gender. There was a positive bias towards boys; results demonstrated that boys were 5.2% more likely (significant at $p < .001$) to be judged relatively highly at maths than girls by their teachers, even when test scores were the same.

This finding was in line with earlier work by Doherty and Hier (1988), who used the Vernon and Miller Graded Mathematics Test and compared results with teacher perceptions. They too found that even when academic competence (the score on the test) was held constant, boys received a "significantly higher predicted mean score" (Doherty and Hier, 1988, p. 342) from their teacher.

Interestingly, in the Campbell (2015) study, when analysing only the results of pupils who were rated below average, boys were 2% more likely (significant at $p < 0.1$) to be judged below average at maths when compared to girls with the same score. This underestimating of boys in maths was also found by Hartas (2018). Results showed that boys were 39% more likely to be placed in the middle/lower set for maths and that gender (along with other teacher perceptions and characteristics) was a sound predictor of set position. Hartas (2018) considers this to be in line with gender perceptions in wider society and considers that it may be due to an attitude gap where boys are less attentive in the classroom than girls.

Whilst these different studies appear contradictory, they all demonstrate that gender plays a role in teacher judgements of mathematical ability, whether over or underinflating it. It may be that wider society and social construction suggest that boys are better at maths (leading to an over inflation when boys demonstrate aptitude) but when boys appear to struggle at maths,

they are judged more harshly because they have fallen so much further than expectations suggest.

A reversal of gendered expectations is found when considering attainment in English. Returning to the Campbell (2015) study, results showed that, in reading, boys were 4% less likely (significant at $p < .05$) to be judged above average and 5% more likely (significant at $p < .001$) to be judged below average by their teacher even when their test score was identical to that of a girl. In a further analysis in which income, SEND status, ethnic origin, and EAL status were held constant (as well as test score) boys were still 3% less likely than girls to be judged above average at reading. Similarly, when considering reading scores, Doherty and Hier (1988) found that when girls and boys were both less favourably perceived (as measured by a personal rating scale), girls received a significantly higher mean predicted reading score than boys. This meant that negatively perceived boys were additionally disadvantaged by lower expectations of their reading ability. In a further example of disadvantage for negatively perceived boys, results showed that there was a wider discrepancy in predicted reading scores between them and favourably perceived boys, than there was between negatively and favourably perceived girls.

Hamilton and Jones (2016) interviewed two teachers from a state primary school in Wales and found evidence of differing expectations. One teacher (recently qualified) felt that literacy in boys was an area of real concern and had been for a long time. She explained that 78% of the boys in her class were in her low ability group for literacy and argued that this was because of their distinctly different learning style. However, the other teacher (a more experienced practitioner) did not refer to different subjects or learning styles in her reflections. The researchers then considered a previous study (Thompson, 2011) which argues that girls being better at reading and writing is now part of contemporary educational belief. Indeed boys being "disinclined towards writing" was a belief found in the interviews conducted by Jones and Myhill (2010, p. 553). They noted that a belief in girls being better writers was expressed by more than twenty percent of participants. This was the only comment to have this highest level of consistency. Other respondents commented that girls were better at spelling, grammar and

English skills. More than four respondents stated that boys are not keen readers, and more than one felt that boys do not like writing.

In further evidence of teachers' perceptions of boys and English, Hartas (2018) found that boys were 50% more likely to be placed in the middle or lowest set for English, although it should be noted that teacher perceptions of behaviour, school attitude, post-16 education goals, children's decision making, and family background were also strong predictors of set position.

Stage Two: Teachers' Observable Behaviour. Two papers specifically discussed differing expectations of male and female pupils being demonstrated through observable teacher behaviours.

Jones and Myhill (2010), observed that, whilst no child in a class of 30 was invited to answer a question very often, underachieving girls were consistently less likely to be invited whereas underachieving boys were consistently more likely to be selected to answer. They suggest that this is a disciplinary strategy to include underachieving boys in the lesson, but which is not as necessary for underachieving girls. Girls are considered less likely to need this disciplinary technique.

During interviews, Hamilton and Jones (2016) explored with teachers how and why they behaved in certain ways towards girls and boys in their classroom. One (who had been teaching for 26 years) felt that they engaged the children in the same way, regardless of gender, but that they tried to make lessons interesting for boys with topic choices being shaped by the cohort each year. The second teacher (who had been teaching three years) explained that they were very aware of different learning styles between girls and boys having been taught this at university. They explained how they had brought in a Snow-White play set to help engage the boys with the story who had found it too feminine. She explained that she was very conscious of boys being kinaesthetic learners and therefore incorporated this belief into her planning alongside teaching boys in smaller groups.

These suggest that teacher behaviours are influenced by beliefs and expectations around gender and that some of these are within teachers' conscious and chosen control. What is not as clear, is whether these are widespread beliefs and behaviours, as sample sizes and interview numbers are very small. The comment about universities teaching learning styles may suggest that some gender difference and learning styles theory (N. D. Fleming, 1995) has been taught in initial teacher training courses despite criticism of learning styles theory being widespread. Indeed, learning styles have been criticised and called a neuromyth for many years (Geake, 2008; Newton, 2015; Riener & Willingham, 2010).

These papers also explored expectations of teachers with a specific focus on what was 'typical' for girls and boys. Jones and Myhill (2010) tallied the number of positive and negative comments made by teachers about girls and boys during their interviews. Girls received more positive comments than negative whilst boys received the opposite. This tally also demonstrated that teachers applied gendered expectations to their analyses of behaviour; the most frequent comments were negative statements about boys' behaviours, followed by positive comments about girls' behaviour. The teachers in the study indicated that girls are considered compliant, quiet and sensible. They want to please, are passive, and fit into a classroom environment. One teacher articulated her belief that, "the girls, I'm sorry to say, are more sensible" (Jones and Myhill, 2010, p. 556). These beliefs around high achieving and bright but quiet girls, may lead to particular teacher behaviours noted in the study such as encouragement and prompting. In contrast, boys are considered as active, "outgoing, needing a challenge or disruptive" (Jones and Myhill, 2010, p. 553). This was also reflected in vignette two (Hamilton & Jones, 2016) in which the teacher reflects that boys are not able to sit still and concentrate for as long as girls.

One teacher in the Jones and Myhill (2010) study was observed to be more tolerant of her female pupils, offering more praise and positive comments to girls than she did to boys (16 to three respectively). The teacher commented to the class that she had lovely girls and that her girls were fantastic.

When considering how best to teach boys, clear instructions, guidance, and pushing, were popular comments (Jones & Myhill, 2010), smaller groups and extra support (Hamilton & Jones, 2016), along with caring but objective teachers (Doherty & Hier, 1988).

Strength of the Literature

Data Gathering Considerations.

Sampling. Across the quantitative studies, the quality assurance process indicated strengths in sample sizes and recruitment with all studies noted as using appropriate participants and recruitment strategies. (A summary of the quality assurance process is included in Appendices B and C). Large data sets were accessed through research such as the Millennium Cohort Study (Campbell, 2015; Hartas, 2018), the National Foundation for Educational Research (NFER) and the Evaluation of National Curriculum Assessment (ENCA) (Plewis, 1997) and project JUDE (Jones & Myhill, 2010). A slightly smaller study (Cline & Ertubey, 1997) achieved a sample of 523 teachers across 79 primary schools. However, in terms of sample size, the two studies with the smallest numbers of participants were Doherty and Hier (1988) with five teachers and 113 pupils from one school and the pilot, qualitative study by Hamilton and Jones (2016) with just 13 children and two teachers. This indicates that generalisations across the country, whilst potentially possible from some studies, must also be considered with caution when reflecting on this set of studies as a whole. Indeed, none of the quantitative studies discuss power calculations for determining sample sizes.

A disadvantage of using a data set collected by others is that it is difficult to speak to any bias within the sample, or as noted by Plewis (1997), be sure of the exact definitions used when coding and categorising data. Quality assurance processes indicated that there was a lack of focus on whether actions were taken to reduce bias within the participant sample in four of the studies. Of the three remaining studies, two studies (Campbell, 2015; Hartas, 2018) stated that the data were weighed because of recognition that there was over or under representation of families with socio-economic deprivation and some attrition, demonstrating that attention was paid to recruitment and sampling bias. The third study (Jones & Myhill, 2010) recognised that the

geographic area for the research was mostly white and middle class but argued that the schools recruited did offer a true picture of intake demographics. With the exception of these acknowledgements, participant recruitment was a weakness across the studies due to self-selecting or purposive sampling, or lack of information. Therefore, any generalisations should be undertaken with caution.

Standardised Testing. Three studies (Campbell, 2015; Doherty & Hier, 1988; Plewis, 1997) used standardised tests as a measure of children's academic competence against which to compare teacher assessments and judgements. Plewis (1997) states that the research is not suggesting one is a more valid indicator than the other, simply that they are both trying to access and measure an underlying ability. Indeed, in the discussion section of the paper, there is reference to criticism of the argument that SATS provide a more accurate picture of a child rather than using information from a teacher who has known the pupil over a period of time. However, later in the research (addressing the third research question) precedence is given to the "supposedly more objective SAT level, rather than to the more subjective TA level" (Plewis, 1997, p. 239) which enables the exploration of bias in teacher judgements.

A further shortcoming of this standardised testing approach to researching bias (and highlighted by the quality assurance consideration of appropriate measurement instrumentation), is that the Vernon and Miller Graded Arithmetic-Mathematics Test would have been the 1976 version when Doherty and Hier (1988) conducted their research, and the Schonell Silent reading test appears to be from 1971. A criticism of standardised tests is that they are not always culturally relevant and this may be an example of the difficulty of relying on such assessments for a 'true' picture.

Campbell (2015) argues that standardised assessments can contain significant errors and patterns of bias and inequality in teachers' expectations of pupils (which are based on gender). However, she further argues that teacher judgements for research purposes need to be made independently of those required by the education system which are used for multiple purposes including judgements of the teacher themselves. Inherent in this system is a vulnerability to bias

through either teaching to the test to increase scores but knowing the child does not fully understand, or manipulating marks to within certain parameters. Therefore, in Campell's research, separate 'research only' assessments were required of teachers about their pupils. Furthermore, the researcher ensured that there was minimal time lag (mean 3.8 months) between the cognitive test of the pupil and the assessment made by the teacher. Whilst this may address one possible source of bias inherent in teacher judgements, this still represents a whole academic term and additional learning or changes may have occurred. Furthermore, it does not eliminate the presumption that standardised tests offer the 'true' picture.

Definitions. Four studies were particularly rigorous in ensuring that definitions were clear regarding different aspects of the research. Plewis (1997) made explicit in his study that references to 'teacher expectations' meant he was using teacher assessments of individual pupils as a proxy measure for expectations linked specifically to that child. Jones and Myhill (2010) found that the term 'underachievement' was problematic because it could mean a pupil who was a low achiever but achieving their best, or a pupil who was currently under achieving but teachers believed they were capable of much more. Following this confusion in the pilot study, a clearer definition was given in the full study. In the study asking teachers to rate children as above or below average (Campbell, 2015), a clear instruction was given that this judgement was to be made in reference to other children of the same age. This ensured that teachers made appropriate comparisons in line with the research aims and avoided making comparisons between children of differing ages but a similar learning profile. Finally, the study in which teachers were asked to reflect on seriousness of need (Cline & Ertubey, 1997) designed a clear hierarchy of levels to ensure teachers were fully aware of the researchers' views of possible steps within a graduated response. This ensured that there was a shared understanding of the concept 'seriousness.' Of the three remaining studies, Hamilton & Jones (2016, p. 245) note that they gave "careful consideration" to the questions used in the interviews to ensure they would be understood by the 6 and 7 year old boys. The other two studies did not comment on whether specific definitions of words or concepts were considered by the researchers at the outset or

found to be a strength or limitation or required clarification for participants throughout the research.

Qualitative Reflections. Of the two qualitative studies, Hamilton and Jones (2016) attempted to gain a richer understanding of the issues through greater consideration of the relationship between researcher and participant. One of the researchers took a job as a classroom assistant in order to have greater contact with participants and build rapport with the boys whom they would later interview. Whilst this may certainly have increased openness and affinity in the relationship, and a number of ethical considerations are noted, there is no discussion of the ethics surrounding whether participants knew in advance that they were a researcher. This is highlighted as a weakness by the quality assurance process. Similarly, the other qualitative study (Jones & Myhill, 2010) did not outline ethical considerations or the role of the relationship between researcher and participant. Where this study offered greater clarity was in a detailed information section regarding the structure of the interviews and the types of questions asked, therefore enabling a more thorough consideration of potential bias.

Data Analysis Considerations. A further strength seen across the quantitative studies following the quality assurance process was in rigorous and appropriate data analysis. The one exception was in the analysis of the interviews conducted as part of the Jones and Myhill (2010) qualitative study. In this instance, a frequency tally for similar comments heard across the 36 teachers was used rather than a more analytical approach. This is of concern because exploring, understanding and critiquing of the methodology is not possible without a clear explanation of the process of analysis, whether coding took place, how particular comments were selected and knowledge of the researchers' role and philosophical paradigm.

Four papers were especially careful to run analyses including, excluding or considering a number of other variables in order to pinpoint or compare the impact of gender. The slightly older studies (Doherty & Hier, 1988; Plewis, 1997) analysed the impact of ethnic group, social class and personal ratings between them. More importantly they both controlled for academic competence in order to consider the amount of variance explained by gender. This was also seen in the newer,

larger studies (Campbell, 2015; Hartas, 2018) in which many variables were analysed between them. These included income, SEN, five different ethnicities, EAL, teacher perceptions (of child behaviour, attitudes, post-16 education goals and parental interest in school), child psychosocial and cognitive characteristics (including self-esteem and wellbeing), and child family background (including subjective experience of SES). This more robust approach to controlling for other contributing factors enables greater confidence in the findings.

A limitation in five of the studies was the lack of follow up, with Cline and Ertubey (1997) being the only study to follow up the impact of sharing data on teacher judgements with participating schools. One paper (Doherty & Hier, 1988) was a follow up and replication study from previous work. None of the other researchers followed up their studies which limits understanding of the impact of teacher awareness upon future expectations.

Discussion

Summary and Overview

This review builds upon the systematic review by Wang, Rubie-Davies and Meissel (2018) into teacher expectation literature and follows up their recommendation of focusing on a sub theme from their results. The review explores the influence of pupil gender on the judgements of teachers and considers whether pupil gender is a factor in teachers' decisions and expectations of pupil academic outcomes. It takes a closer and more nuanced look at research contexts, methodologies and controlled variables. Furthermore, it explores potential causes for differing findings in results.

Existing literature and wider research in this field presents a mixed finding as to whether there is a gender bias within teacher expectations. Hecht and Greenfield (2002) researched 170 first and third grade children in a multi-ethnic urban city in the United States and found that gender accounted for small amounts of variability in teacher judgements but suggest that it was not a substantial contribution. A more recent study by Jabůrek et al. (2022) in the Czech Republic (pupils = 223 mean age 10.34) found no significant difference in teacher expectations between

genders. Within this review, only one study did not find a significant gender bias in expectations. This was Cline and Ertubey (1997). Their methodology used fictional characters described in vignettes with minimal data and no externalising behaviours described.

In contrast to Cline & Ertubey (1997), one study in this review did find that girls were more likely to receive a higher teacher assessed grade than boys across three subjects: English, maths and science when compared to a standardised test (Plewis, 1997). Further evidence suggests that expectations of teachers may differ for boys and girls in line with different subject areas. Two further studies within this review, Campbell (2015) and Doherty and Hier (1988) both found that teachers were more likely to judge boys more favourably at maths even when they achieved the same test scores as girls. This is consistent with a study by Rubie-Davies and Peterson (2016) from New Zealand (pupils = 2234 mean age 11.6) in which teachers were twice as likely to overestimate mathematical achievement on standardised tests in boys.

A summary of the past 50 years of research into teacher expectation effects (Timmermans et al., 2018) argues that methodological issues should be a key concern as research moves forward. Past studies often differ in how expectations are measured, whether certain characteristics should be controlled and definitions around whether 'expectation' refers to the present or the future. This study adds an extra dimension to the methodological considerations in suggesting that a vignette approach may lead to different conclusions than a standardised testing approach. Aguinis & Bradley (2014, p. 353) argue that vignettes can be helpful in understanding "implicit decision-making processes", and can provide insight into non-observable phenomena. The authors suggest that vignettes also offer researchers the opportunity to include relevant aspects of their experimental focus whilst ensuring that potentially confounding factors are left out. Furthermore, Baudson & Preckel (2013) emphasise that asking teachers to consider real individuals that they have taught can lead to a wide range of pupils' characteristics being brought into reflections and decision making, which may not be an intended part of the research. Aguinis & Bradley (2014) suggest that all vignettes used in a study should be published and available to the reader, as indeed they are in the Cline & Ertubey (1997) research.

Cline & Ertubey (1997, p. 449) state that they provided participants with “rounded” vignettes and utilised a “realistic” and “contextualised” rating score. It was when using a vignette methodology and these more ecological research materials, that they argue the impact of gender on teachers’ judgements disappeared. However, Finch (1987) warns that researchers should not presume that people will act in the same way they have stated in hypothetical vignettes. Similarly, vignettes can oversimplify complex real-world situations and remove any emotional responses. This can lead to idealistic responses or answers which are not then generalisable to real life (Schwappach et al., 2013). In the vignette-based research examined in this systematic review, participating schools were self-selecting on a first-come-first-served basis which may have led to bias in the sample (e.g., those who felt confident in their understanding of SEND) further limiting the generalisability of the study. This potential drawback was also recognised by the authors who comment on the geographical context and pupil ages in the study being limited.

Holder and Kessels (2017) used a vignette approach to examine student teachers’ expectations of third-grade pupils in maths, in Germany. They provided participants with identical vignettes but told them they had been written based on a child’s portfolio of work (subjective) or written based on a child’s performance on national tests (objective). In each condition, half were attributed to a male pupil and half to a female pupil. In the subjective condition, participants were asked to rate the pupil’s performance on a linear scale (very poor to very good), and this was later converted to a numerical score by the researchers based on measurement along the line. In the objective condition, participants were asked to estimate the number of points the child received on the national test. Results demonstrated that in the objective condition, females received lower scores, thus revealing a bias. However, in the subjective condition, there was no difference between male and female scores which the authors argue is because of a predicted ‘standard shift.’ Holder and Kessels (2017) explanation is that teachers shift their standards for certain groups. This means they make their judgements relative to others in the group (e.g., a boy's attainment in maths is judged in comparison to other boys' attainment in maths and girls in comparison to girls). Therefore any bias is disguised when using a subjective scale for rating e.g.,

class results demonstrating equal numbers of boys and girsl were rated 'very good' could disguise the fact they needed to achieve different standards to earn this label. This could be the case in the vignettes. However, objective scales can expose bias because they “maintain a constant meaning regardless of who is judging or who is being judged” (Holder and Kessels, 2017, p. 475).

This suggests that methodologies may have an impact on identifying teacher biases within their expectations, with use of standardised data more able to demonstrate any such biases. When considering gender and academic attainment, this would mean that teacher expectation scores can be directly compared to a pupils’ standardised test score thus revealing any bias. Any potential shifting of standards based on gender would become apparent.

Quantitative research using objective scales both in this review (Campbell, 2015; Doherty & Hier, 1988) and in the wider field found that teachers hold higher expectations for girls than boys in English. A longitudinal study by Hinnant, O’Brien and Ghazarian (2009), comprising of nearly 3000 pupils from 10 locations in the United States, compared teacher perceptions of attainment to standardised data, and found that girls were rated as higher, and boys as lower, than their scores indicated. Qualitative studies from this review (Hamilton & Jones, 2016; Jones & Myhill, 2010) spoke to teacher beliefs that boys have a different learning style and girls are better writers, whilst boys do not like writing. Ready and Wright (2011) conducted a longitudinal study analysing North American kindergarten children (n = 9493) and argue that there are “stark gender differences” (p. 346) when analysing the children described as having strong literacy skills. They found that 55.1% of pupils in this category were female. Whilst the authors report a less than 0.001 p value, they acknowledge that this statistical significance is partly due to the large sample size. This means that the clinical significance or real-world relevance is reduced, so whilst there may be “systematic error or bias in teacher perceptions” (p. 354), the impact is limited. Further analysis found that the overestimation of girls’ literacy skills remains across the whole of kindergarten year, even as teachers get to know their pupils better. Consistent with these findings is a longitudinal, New Zealand study (Meissel et al., 2017) analysing standardised attainment scores and overall teacher judgements in reading (n=4771) and writing (n=11,765). Data from

pupils across the whole school age range (years 1 – 13) found that when standardised attainment score is the same, females receive significantly higher overall teacher judgements than males. It appears that, when using a standardised testing methodology, differences in teacher expectations are apparent and significant for different subjects with boys being favoured for maths and girls for English. It appears then that large scale, longitudinal, international studies are demonstrating similar findings to this UK based review.

The original paper that led to this review (Wang et al., 2018) began to consider the complexity of the process of forming teacher expectations, suggesting that demographics, perceived pupil characteristics and overt classroom pupil behaviours all play a role. Six of the studies in this this review, began to consider the relationship between gender and other possible sources of bias. Wider literature has also emphasised the importance of this as an approach to research in this area, for example Timmermans et al. (2016) suggest that differences in expectations between the genders is present in behavioural expectations as well as achievement predictions. Motivation and work ethic has also been found to interact with gender in the formation of teacher expectations (Gentrup & Rjosk, 2018) and one study argued that teachers only expect girls to perform as well as boys in maths when the girls are perceived as extremely hard working (more so than the boys) and with an eagerness to learn (Gentrup et al., 2018). This difference in behavioural expectations was certainly the case in the 'typicality' and 'atypicality' discussions of the different genders with 40 teachers (Jones & Myhill, 2010) and explored in terms of a 'personal rating scale' by Doherty and Hier, (1988) in which they too found complex interactions between teacher perceptions on the rating scale and the gender of the pupil, noting that boys could be subject to double negative perceptions.

Studies in this review highlighted ethnicity as a possible factor interacting with gender when teachers form expectations. This potential interaction has also been explored in the wider literature. Hinnant, O'Brien and Ghazarian (2009) found that by the fifth grade, ethnicity as well as gender was a significant factor in teacher expectations. A later study (Riegle-Crumb & Humphries, 2012) looked specifically at intersectionality between gender and ethnicity. Using data from the

Educational Longitudinal Study of 2002 (10,486 students in 10th grade in the United States) they investigated teachers' potential bias in expectations of maths ability. They found evidence of a biased belief that maths is easier for white males than for white females but that ethnic minority students are not rated less favourably than white males. Indeed, teachers of advanced maths are less likely to suggest that courses are too difficult for black females than they are for white males. Whilst the paper goes on to consider reasons for this finding, the authors suggest further qualitative studies in the area would be helpful.

All studies in this review considered gender as binary categories. However, Monro (2005) argues that as there are people "who can be interpreted to be other than male or female: intersexes, transsexuals in transition, and androgynes" (Monro, 2005, p. 3), there is a need to replace the binary understanding of gender with a spectrum conceptualisation. This entails the breaking down of distinct male and female categories and the embracing of a more gender diverse and gender fluid approach which includes models of gender based on physical, social and psychological aspects as well as changes in social structures and further academic research. This notion would be considered a gender-pluralist approach. In a survey of gender diverse adolescents, two-thirds identified themselves as non-binary rather than transgender (Baum et al., 2012) which suggests they identify as both, neither, or somewhere between male and female. It may also suggest fluidity whereby their gender identify fluctuates over time (Diamond, 2020). This again questions the binary distinction between genders.

If gender is more complex than a simple binary approach, the question arises as to how to measure or assess this. Gülgöz, Edwards and Olson (2022) argue that there is still no agreement as to how gender is best measured. Some studies consider how well-matched children's perceptions of the self are to others of the same gender. Some use a continuum approach to probe children's feelings around their own sense of gender, whilst others utilise several measures to build up a picture of gender as a multidimensional construct. In their own work, they develop a single-item continuous measure of gender with totally male and totally female at either end.

It is unknown what views teachers within the studies in this review hold regarding the concept of gender identity, but as understanding of gender as a spectrum rather than binary develops, broad assumptions around subject specific abilities and behavioural traits linked to each gender need to be further questioned. Teacher expectations of academic attainment or behaviour traits of pupils may be influenced by their understanding and beliefs around gender. Future research will need to be aware of this, of how gender definitions are considered in society at large, and how children are identifying themselves in relation to gender. It will be important to consider how studies can accurately capture and account for these factors.

Implications for Future Research

As suggested above, further research into intersectionality effects would be welcome in this area and widen the debate beyond the potential standalone effects of gender. Whilst statistical models offer researchers the opportunity to control for certain variables such as ethnicity or views of behaviour whilst exploring the impact of gender, perhaps there is a more pressing need to consider cumulative effects alongside this.

It would also be beneficial for future research to explore potential between-teacher or between-school effects, as recognised by Plewis (1997) when he argues that results are averaged across teachers and may not be as unambiguous as this. Cline and Ertubey (1997) did explore teachers' length of experience, gender, current responsibilities and additional training and found no such between-teacher effects but this was a small UK study and would need to be replicated on a wider scale. Further exploration of between-teacher effects found that teachers who provided highly differentiated work for low and high achievers were more stable in the expectations of pupils. Those offering less differentiation were more likely to modify their expectations over time (Kuklinski & Weinstein, 2000) perhaps because they were able to make more direct comparisons with others in the class or because there was no 'ceiling' on the opportunities provided to the lower achievers.

Between-school effects were considered by Agirdag (2018) with a particular focus on the area of socioeconomic composition and the link with teacher beliefs. He found that collectively,

teacher beliefs were affected by school characteristics (e.g., SES) and suggested that teachers need to fully understand not only how their beliefs and expectations can impact individual pupils but how those beliefs and expectations can be shaped by whole school characteristics and staff culture. This supports the suggestion that further research into between-school effects is an important avenue in the field of teacher expectation research.

None of the studies in the review took a longitudinal approach. However, wider literature in the field is beginning to look at stability or adaptability of teacher expectations over time, with studies finding that ordinal rankings and individual assessments of primary aged pupils were very similar across the academic year (Kuklinski & Weinstein, 2000; Rubie-Davies et al., 2018). Contrastingly, Timmermans et al. (2021) found that teachers would adapt their expectations for pupils across the year based on their individual achievements (although this did not always impact the rank ordering within the class) and Hao et al. (2022) suggested a more complex picture with older pupils (grades 11 and 12), whereby teachers were willing, in the initial months of the year, to alter their early expectations, but these then remained stable for about a year. Expectations then fluctuated again near the end of the second year which the authors suggest may be due to contextual factors such as teacher stress or pressure to do well in exams changing pupils' behaviour and achievement. These studies are particularly important as they have implications for training and encouraging self-reflection within teachers. Furthermore, it may be that teachers believe and expect perceived gender differences to remain constant, or it may be that teachers perceive there is a certain level of malleableness before or after a certain age. If this is the case, it would be helpful to research teachers' beliefs about the permanence of any perceived gender differences.

Implications for Practice

While the literature in six out of seven studies in this review does find differences in teacher expectations relating to pupil gender, the two qualitative articles make particular reference to the importance of not treating girls and boys as homogenous groups. Jones and Myhill, (2010) argue that it is underachievers who have more in common with each other than

other pupils of the same gender. The implication for practice here is that overarching strategies designed to support one gender or the other may not be as effective as those with a more nuanced, individualised approach.

Campbell, (2015) suggests that an implication for practice is that whilst national resources and policy initiatives need to be focused on addressing teachers' inaccurate expectations which can impact a pupil's learning and attainment, they can also have the exact the opposite effect. National initiatives, such as pupil premium or OFSTED required analysis, can reinforce beliefs that certain characteristics inevitably lead to lower attainment and thus reinforce the exact expectation it is designed to combat.

A further area in which to consider additional training and awareness may be initial teacher training. In particular, where this training consists of significant time within one or two placement schools, trainee teachers may be more susceptible to current norms within a workplace rather than critical pedagogy from a university. In these circumstances, access to high quality research with time for analysis and reflection would be an important aspect to include from the very beginning of a teacher's career. Moreover, regardless of type and location of training, initial teacher training programmes should include how teachers can seek to avoid the potential negative effects of low expectations (Timmermans et al., 2018).

Statistics suggest that girls do not seek careers in STEM fields in such as great a number as boys (Lazarides & Lauermann, 2019; Rubie-Davies & Peterson, 2016) and one hypothesis around this is that the expectations of teachers are lower for girls from an early age, which leads to fewer girls seeking to study courses such as advanced maths (Robinson-Cimpian et al., 2014). Alongside initial teacher training courses, EPs may be well placed to ensure continued awareness within the individual school environment through access to high quality research, inclusion of reflections on expectations when delivering training courses, promoting awareness when involved in systemic work, and an individualised approach when supporting specific pupils.

Whilst there is clearly an ongoing need for teachers to be aware of their expectations for pupils which may have some basis in beliefs around gender, Doherty and Hier (1988) offer

reassurance that the amount of variance in teacher judgements that was explained by academic ability was eight times larger than the variance explained through their personal rating scale. This suggests that teachers' predictions are "fairly accurate" (p. 346) and based on pupil academic competence. However Timmermans, de Boer and van der Werf (2016) argue that pupil performance only explained 80% of variance in their study and that the 3% explained by teacher perceptions of pupil attributes (of which gender was one) is still a "considerable amount" (p. 234). If indeed there is a small amount of variance in teacher expectations, then even a small amount is too much and educational systems continue to need the concentrated efforts to combat this as recommended by Campbell (2015).

Strengths and Limitations

This review is novel in that it only considers UK wide literature thus making it particularly relevant and sensitive to the cultural and gender expectations of a specific location. It benefits from large samples of children and teachers (up to 9610), studies up to 2018 and the inclusion of qualitative studies seeking an insight into what may be contributing to potential gender effects on teacher expectations. It included a search of a grey literature database (ProQuest) to reduce the risk of publication bias and increase the robustness of any findings. However, although articles were found through this method, none met the inclusion criteria for the review.

This review built on the recommendations of Wang, Rubie-Davies and Meissel (2018) through pursuing their proposal of a more detailed analysis of a sub-theme in their paper. It reflects findings in wider literature from across the world that both supports and contradicts the suggestion that teachers hold different expectations of pupils based on their gender. Where this review offers something new is in the suggestion that chosen methodology may be accounting for findings in one direction or the other.

The selection process for papers included in this review was through following clear and recognised guidelines for systematic reviews (Page et al., 2021), and studies were quality assessed using validated and recommended tools. However, whilst this process sought to be robust, it is possible that some relevant studies were not discovered. Only seven studies met the inclusion

criteria which therefore limited the available research from which to base the systematic review. This means it is difficult to make clear assertions in answer to the review question. However, whilst a wider inclusion criterion than UK was considered (e.g., countries with a similar curriculum structure or systemic approach to education) these concepts would need detailed reviews in themselves and may still result in the lack of generalisability to the UK.

Finally, Plewis (1997) articulated an important limitation for all work in this area which is that even if findings all suggest a teacher expectation effect, more work must then be done to investigate how this might translate into actual teacher behaviour in the classroom. In this review, Hamilton and Jones (2016) noted that teachers may change the content of the curriculum or style of delivery, and Jones and Myhill (2010) considered which children were asked questions or deliberately drawn into the lesson. In wider literature, Robinson-Cimpian et al. (2014) begin to discuss the importance of feedback to pupils or how teachers' own subject anxieties may be implicated in transferring expectations to pupils. All of these behaviours may subtly transmit teacher expectations to pupils and therefore any research which increases teacher awareness and understanding of this area offers the opportunity to impact the classroom experience for pupils.

Conclusion

To conclude, this review found evidence of pupil gender influencing teachers' expectations of primary aged pupils in the UK and this is supported by wider literature across countries and ages. However, there is still debate in the wider literature as to whether this is always the case. Further research which considers methodology, gender as a spectrum, subject differences and malleability over time are all still very much needed in this area.

3 - Teacher Goals and the Distribution, Content and Self-perception of Teacher Language Across Contexts

Abstract

The purpose of this study was to explore the nature of, and differences between, teacher language usage in whole class contexts compared with small groups. It also considers teachers' self-perception of their language including any factors which they deem may have impacted their language. In this convergent mixed method empirical study, observations of 12 teachers in their primary classrooms were conducted using the OPTIC. Initial descriptive statistics were calculated and discussed in semi-structured interviews which explored teachers' thinking and attitudes about their language usage. Detailed observation data were later analysed using a repeated measures analysis of variance and showed that there was a significant interaction ($p < .001$) between type of language used by teachers and the context in which they were teaching.

Five analytic themes were generated from the interview data using Thematic Analysis: beliefs and knowledge of pedagogy, having clear boundaries and expectations, individual needs of children, the environment, and reflections about the self. These are discussed in conjunction with the quantitative data.

Implications for practice are discussed with direct reference to the Stages of Change model (Prochaska & Di Clemente, 1982). Helping children learn requires a breadth of thinking that includes adjusting the environment (e.g., making use of small group contexts) adapting language for individual pupils, and the careful use of non-verbal communication. A checklist resource is developed and offered to support teacher reflections on their use of language and to consider the reorienting of strategies to further think about enabling learning and promoting positive environments.

Suggestions for further research include considering which factors can support teachers in the accuracy of their predictions, additional research into how language changes across different contexts, and an update to the literature reviewing positivity studies in the classroom.

Praise can be described as *"any verbal statement or gesture that indicated teacher approval of a desired student behavior [sic] beyond confirmation of correct academic responses"* (Reinke, Lewis-Palmer and Merrell, 2008, p. 319). In contrast, reprimands were defined in the same study as "*verbal comments or gestures made by the teacher indicating disapproval of student behavior [sic]"* (Reinke, Lewis-Palmer and Merrell, 2008, p. 319).

Internationally, several recent studies have shown that teachers use more disapproval and reprimands than positive comments when speaking to their class (Floress et al., 2022; Reinke et al., 2013; Sulla et al., 2019). Drake & Nelson (2021) report a mixed picture in their systematic review. In a subset of their studies (those which measured both praise and reprimands), they found three with higher rates of praise and four with higher rates of reprimand, although their search criteria had not included reprimands specifically so this may limit the conclusions in this area. However, a 30 year review of studies across seven countries and dating from 1970 to 2000 (Beaman & Wheldall, 2000) found a trend of teachers using more approval than disapproval overall. In support of using high rates of approval in classrooms, as was found in the review, having a high ratio of praise-to-reprimand comments has been found to correlate with the time pupils spend engaged and on-task (Caldarella et al., 2020; Nafpaktitis et al., 1985) which in turn may be linked to better pupil outcomes (Fisher et al., 2015). In order to further explore the definitions and impact of positive language, Royer et al. (2019) conducted a systematic review which focused on studies which had categorised positive comments into 'general praise' and 'behaviour specific praise.' They concluded that behaviour specific praise could be considered an evidence-based practice and that it was an effective way to reduce unwanted behaviours and increase appropriate behaviours in the classroom. These findings support the use of positive language in the classroom, but Floress et al. (2021) articulate a difficulty, arguing that there is a lack of clarity as to the level of positive teacher language occurring in classrooms without specific interventions (known as natural rates (Drake & Nelson, 2021)) and whether natural rates meet certain recommended ratios. This suggests that any broad-based professional development

recommendations should consider the accuracy and effectiveness of baseline levels before advocating for change.

The Education Endowment Foundation guidance report on feedback to improve pupil learning (Collin & Quigley, 2021) states that feedback can be about a task, a process, or a pupil's self-regulation. The aim of feedback should be to improve pupil learning outcomes. The report recognises that there are a variety of methods and approaches for delivering feedback. However, it was teachers' verbal behaviours that were specifically considered by Apter, Arnold and Swinson (2010). They scrutinised teacher language for 'positive' and 'negative' comments whilst also considering whether these were regarding academic work or pupil behaviour. The authors found that teachers used significantly more positive verbal comments in response to pupil behaviour than had been reported in earlier studies (Harrop & Swinson, 2000; Merrett & Wheldall, 1986). A review of UK classroom observations (Apter et al., 2020) found that, in both primary and secondary schools, the most common use of teacher language was to make positive comments about academic tasks, followed by negative comments about behaviour. Positive comments regarding behaviour were less frequent and lastly was the use of negative comments regarding academic work. The aforementioned study by Sulla, Armenia and Rollo (2019) shows a similar pattern of higher rates of approval for academic work and of criticism of social behaviours.

The sequence of actions described by the Initiation-Response-Feedback (IRF) model (Sinclair & Coulthard, 1975) is frequently seen within classrooms (Zare-ee & Hejazi, 2019), and describes the three-part structure of many teacher and pupil interactions. Whilst some researchers are critical of its influence on these interactions (Fagan, 2018), arguing that it impedes learning and can close down conversations (Waring, 2008), others argue that it is "quite nicely designed" with a "built-in repair structure" (Newman et al., 1989, p.127). Therefore, whilst some teachers and researchers may see certain comments as 'negative' or 'criticism', there is also the possibility that some will argue these types of verbal behaviour have a benefit of their own.

Apter, Sulla and Swinson (2020) also demonstrated some variance across contexts, with secondary school teachers making fewer behaviour comments (either positive or negative) than

primary school teachers, and staff in maths departments giving significantly less feedback of any kind compared to the English department. Additionally, there are differences in the verbal behaviour of teachers towards different genders in primary school with boys receiving more instructions, approval and disapproval than girls (Harrop & Swinson, 2011). A systematic literature review of rates and types of teacher praise (Jenkins & Floress, 2015) found that instructional activity (e.g., whole class instruction, small group instruction, independent work, transition time) was not specified in studies, leaving the authors unclear as to whether there were differences in use or type of praise in different contexts. Research and greater understanding in this area could impact the practice of current teachers as well as guiding initial teacher training programmes (Floress, Zoder-Martell, et al., 2021).

Group work itself has been the focus of much research, including the SPRinG project (Social Pedagogic Research into Group-work) which aimed to address the disparity between the potential of group work to make a positive difference to learning and interactions in the classroom, and the amount of group work taking place in schools (Blatchford et al., 2006). The SPRinG project "stressed collaborative and autonomous learning" (Blatchford et al., 2006, p. 751) and supported teachers to use a more scaffolded approach in their teaching style, rather than instructive. The follow up handbook for teachers and practitioners (Baines, 2017) stresses that, in effective group work, children should be co-learners and that teachers should be aware of the difference between children working as a group or in a group. Teacher language behaviours should be less about controlling the group or task completion, and more about monitoring, guiding pupils and teaching group working skills. A Scottish extension to the SPRinG study (Howe et al., 2007), researching 24 science classes of 10-12 year olds, found that group work played an important role in pupil understanding. These group work opportunities had been supported firstly by teacher professional development in implementing effective group work and secondly through the teacher delivering group-skills education to the pupils before any curriculum content. In a further study of secondary science and English classes, researchers considered (among a variety of factors) differences in teaching approach (group work or whole class) and differences in teaching

style during distinctive stages of the lesson (introduction and conclusion). Researchers found that "teachers adopt a consistent approach throughout" (MacQuarrie et al., 2012, p. 539), utilising similar approaches to their introductions and conclusions in both group and whole class contexts. Contrastingly, pupils adapted their behaviour and dialogue depending on context which, whilst suggesting that pupils are able to interact successfully in small groups, still needs further exploration to ascertain whether "group work is as productive as pupil behaviour suggests" (MacQuarrie et al., 2012, p. 540). In concluding, the authors argue that teachers do not appreciate the potential value of modifying their language or strategies when working with pupils in small groups.

Managing a class is not a simple task and there are many decisions to be made within the complex classroom environment: "Teachers can be engaged in 1,000 interactions a day, sometimes more" (Watkins & Wagner, 2000, p. 54) and Good et al. (2018) argue that teachers have to deal with rapid and ambiguous events extremely quickly. Bailey et al. (1983) found that teacher ratings of pupils' on task, disruptive and interaction behaviours did not correlate well with observational data on social behaviour within the classroom. Although this may have been due to the variability in interpretation of what is considered inappropriate behaviour, it leads to questions as to whether teachers are accurate in their perceptions of events in the classroom when making the many decisions necessary. Similarly, teachers may not be accurate when reflecting on their own behaviour and language within the classroom. Kim and Stormont (2016) found that teachers in an Early Years setting reported using more praise than reprimands during a two-hour period which included transition times, free play and group times. These recollections were not in line with observations of the lesson. Contrastingly, Floress et al. (2022) found an overall significant correlation between perceived use of praise and observational data when reflecting back on a 20 minute lecture-based lesson. Notably, the observation was shorter than the study by Kim & Stormont (2016) and did not include transition times or independent work times. However, the research by Floress et al. (2022) also found that teachers who were not accurate in their perceived use of praise were also not accurate in their perceived use of

reprimand, suggesting that for some teachers, reflections on use of language were difficult. This suggests that any professional development around teacher language use needs to carefully consider the most helpful use of objective and subjective data and recognise that teachers may not have full or accurate insight or recall of their many and varied interactions.

Priorities of teachers in decision making and use of language must also be considered. Pirskanen et al. (2019) found that teachers of young children prioritised happiness and excitement in school alongside the teaching of emotion management. Academic skills were felt to be of secondary importance, developing later in the year. Focusing on behaviour praise may also lead to a more predictable, friendly environment which promotes a sense of security for pupils (Spilt et al., 2016) and therefore may impact the choices and priorities of different teachers. Furthermore, teachers may also have self-directed goals which influence their classroom choices; a positive learning environment at any age may be linked with overall teacher well-being (Drake & Nelson, 2021). It is therefore important not just to gather statistical information about the type and quantity of language teachers use, but also what role their attitude, goals and intentions may have in these decisions.

Research Questions

This research sets out to answer four questions.

RQ1: How is teacher language distributed across four categories (positive academic comments, positive social and behavioural comments, redirective academic comments, redirective social and behavioural comments)?

RQ2: What is the distribution of teacher language in a whole class context and how does it change when working in a small group?

RQ3: How accurate are teacher perceptions of the language they use within the classroom?

RQ4: What do teachers say impacts their use of language in the classroom?

Hypotheses

Based on the research team's clinical experience, the hypotheses were each developed with a directional focus. Consideration was given to the teacher having fewer children and therefore a much tighter focus when in a small group context. This in turn could increase opportunities to notice positive behaviours and positive academic work, alongside reducing pupils' opportunities for behaviours which might elicit disapproval. Similarly, teachers might be able to pre-empt learning mistakes for individual pupils, therefore requiring fewer academic redirections.

1. There will be a significantly higher mean rate of positive academic comments in the small group than in the whole class.
2. There will be a significantly higher mean rate of positive social behaviour comments in the small group than in the whole class.
3. There will be a significantly lower mean rate of redirective academic comments in the small group than in the whole class.
4. There will be a significantly lower mean rate of redirective social behaviour comments in the small group than in the whole class.
5. Teachers will be more accurate in their predictions of their use of positive language when reflecting on a small group context.

Method

Pilot Study

A pilot study took place in a small rural school. This consisted of a 60-minute classroom observation followed immediately by an interview with the teacher. During the interview phase, the teacher expressed concerns about the use of the term 'negative language' due to its emotive nature and her belief that teachers have to convey to children when something is wrong in order to support them to correct it. Following the pilot study, two category titles were amended to 'redirective' rather than 'negative.' No further changes were made to the procedure.

Participants

Twelve participants (male = 3, female = 9) were recruited from across six schools covering two local authorities in the South of England. Sample size was determined using the guidelines recommended by Braun & Clarke (2013), who recommend between six and fifteen participants for this type of study, and the importance of a large enough sample to gather data across the primary age range (Reception, Key Stage One, Lower Key Stage Two, Upper Key Stage Two). The same participants took part in both the qualitative and quantitative parts of the study.

These teachers were recruited through previous professional relationships and discussions with Headteachers and schools' link EPs. Headteachers were asked to identify potential participants and any members of staff who expressed an interest in participating were given further information sheets. Each participating school was given a monetary incentive of a £100 contribution to the Parent Teacher Association of the school. This was intended as a token of gratitude for the use of staff time and necessary covering of classes during the interviews. Ethically, this could have placed additional pressure on teachers to participate either from a sense of responsibility to the school or from a headteacher requiring their participation to gain the contribution. However, it was made clear to all teachers that the school would still receive the money if they individually withdrew, in the hope that this would go some way towards mitigating this pressure. The incentive was the same for all regardless of the number of participants.

The social economic status of the schools was measured using Free School Meals (FSM) data and ranged from 3.8% to 41% with a mean of 21.18%. One was city based, three were in a large town and one in a small, rural village.

Teachers' years of experience varied in range as follows: fewer than 5 years (n = 3), 6 - 10 years (n = 2), 11 - 15 years (n = 2), 16 - 20 years (n = 4), more than 20 years (n = 1). The age range of classes spanned the entirety of primary school: Early Years (n = 2), KS1 (n = 3), lower KS2 (n = 4), upper KS2 (n = 3). Class size ranged from 8 to 31 (M = 24.66) and small group size ranged from 3 to 6 (M = 4.83).

Materials

An adapted version of the Observing Pupils and Teachers In Classroom (OPTIC) behaviour observation schedule (Merrett & Wheldall, 1986) was used for recording the types of language used by teachers. (See Appendix E.) Section A of the OPTIC is designed to facilitate the collection of key teacher-behaviour data into four distinct categories. These four categories encompass teachers' use of approval and disapproval of children's academic and social behaviours. Merrett & Wheldall (1986) describe the OPTIC as "easy to master, objective, sensitive and reliable" (p. 60). The authors suggest a complete observation can be undertaken in 30 minutes, although this was extended to 60 minutes for this study to ensure adequate recording time in both the whole class and small group context. Merrett & Wheldall (1986) also offer suggestions for the categorisation of teachers' language which was used as a guide but further developed to include more modern phrases (see Appendix F). In accordance with the original authors, where teachers are giving instructions rather than responding to a pupil, these behaviours are not categorised. Similarly, in this study, where teachers offered guidance for future work or asked generic questions, this was not included in the data. Section B of the OPTIC was not required for this study as it focuses on pupil behaviour.

Consideration was given to the use of the Mixed Interval Class Room Observation (MICRO) as used by Apter (2016). However, the MICRO offered greater opportunities to record pupil behaviours than were necessary and thus the OPTIC was selected.

Design

This was a convergent mixed method design with the initial phase being a classroom observation and the second phase being an interview. Both phases were completed in a school on the same day before repeating the process in a different school. Observations in the class required participants to teach for a whole lesson (usually one hour depending on the school's timetable) and to include a small group focus session as part of the lesson.

Quantitative data were collected using the OPTIC (see Appendix F for categorisation of comments decisions) along with clear information regarding the amount of time spent teaching the whole class and the amount of time teaching a small group. This ensured the dependent variable 'comments per minute' could be calculated during the analysis stage. The proportion of comments of each type was calculated following the observation but in advance of the interview. During the interview, teachers were asked to self-report on their distribution of language across the four categories shown in Tables 2 and 3. Teachers were then shown the data from their observation and asked for reflections including how it matched their recollections and hopes.

Qualitative data were collected through recorded semi-structured interviews in which participants were asked to reflect on and predict their use of language proportions in each of the four categories. See Tables 2 and 3. Interviews were held for 18 to 30 minutes with an average interview time of 24.5 minutes. The added value offered by this mixed method design was that teachers were reflecting on and discussing real, and therefore ecologically valid, classroom situations with their own data. The interviews ensured an opportunity to gather perceptions and interpretations regarding how teachers experience their classrooms and context. This reflects the critical realist approach of the study. Further information on the researcher's background and experiences can be found in the 'Personal interest' and 'Myself as a researcher' sections of chapter 1.

Table 2

Breakdown of Types of Comments in Whole Class Teaching

Whole Class teaching	
Positive academic	Positive social and behavioural
Redirecting academic	Redirecting social and behavioural

Table 3

Breakdown of Types of Comments in Small Group Teaching

Small group teaching	
Positive academic	Positive social and behavioural
Redirecting academic	Redirecting social and behavioural

Participants were also questioned on their hopes and expectations for the lesson and the outcomes they intended when using different types of language. (See Appendix G for Topic Guide)

Quantitative data were analysed using a repeated measures ANOVA and paired samples t-tests. Qualitative data were analysed using a critical realist approach to thematic analysis (Braun & Clarke, 2006, 2020; Fryer, 2022a).

The mixed methods design was chosen based on the researcher's critical realist epistemological stance.

Procedure

University of Southampton's School of Psychology Ethics Committee gave ethical approval for the study (see Appendix H). Once participants had agreed to be observed and interviewed,

they were reassured of their right to withdraw up until interview data was transcribed and that payment to the school would not be impacted by the withdrawal of any individual teacher.

Observations of lessons took place within the normal school day and interviews were held on the same day in a private room. Same-day interviews were intended to give teachers the greatest opportunity for accuracy in their self-reports and in recalling specific interactions with individual pupils. Especially in a primary school, where teachers work with the same pupils all day, it may be harder to recall which interaction took place in which lesson if asked on a subsequent day. Although there was potentially a disadvantage of same-day interviews in that there was less time for the teacher to reflect on choices made during the lesson, the interviews were not rushed to ensure there was ample reflection time within them. Sequencing of questions and topics was carefully considered in advance (Cohen et al., 2018) and the semi- structured nature of the interview ensured that ideas could be followed up and responses explored as required (Bell & Waters, 2018).

Data Analysis

Braun & Clarke (2021a, p. 4) argue that thematic analysis (TA) is a "family of methods" rather than a "delimited methodology" and that it can be utilised within different ontological, epistemological and theoretical frameworks (Braun & Clarke, 2021b). Indeed Fryer (2022a) suggests it is possible for qualitative research to consider causation and that there is a growing literature base to support the development of critical realist methods in this area. Step one of the critical realist approach to thematic analysis involves consideration of the research question itself to ensure it has a causal conceptual basis. Table 4 details the steps of the thematic analysis procedure used.

Table 4

Five Step Critical Realist Approach to Thematic Analysis

Stage of analysis	Analytic procedure employed

1. Development of the research question	• The lead researcher reviewed the literature and identified focus research questions.
2. Familiarisation with the data	• The lead researcher read each transcript carefully whilst listening to the recordings. Any necessary corrections were made.
3. 3.1 Development and application of codes 3.2 Further development of codes	• The lead researcher coded each potentially relevant item in the transcripts using a data-led approach and descriptive codes. (See Appendix I for examples.) • Codes were consolidated where theoretical or other similarities were noted. • Code names were amended if necessary to better describe the data.
3.3 Review of codes	• Steps 3.1 and 3.2 were repeated many times to ensure interpretive and descriptive validity.
4. 4.1 Development of themes 4.2 Review of themes	• Themes were developed which began to answer the research question. • Possible names which encapsulated the contents of the themes were generated. • Themes were reviewed and considered in terms of their ability to explain and account for the

	observations in the classroom and the research question. • During the review process, and following discussion in the research team, some codes and sub-themes were reallocated to different themes.

5. Generation of conclusions and report	• An accessible thematic map was created
	• Consideration and reflection was given to
	  ○ Limitations of the research
	  ○ Implications for practice in the classroom and for educational psychologists
	  ○ Links to wider literature

Member checking of synthesised, analysed data took place in April 2023, immediately following the creation of the thematic map but before the wider considerations in step five. The purpose was to verify whether the themes generated in step four resonated with participants (Birt et al., 2016). The data was presented in an accessible format and participants were asked if they felt their views were represented in the data and whether the themes made sense (Creswell & Miller, 2000). Member checking was positive and confirmatory and did not result in any changes to themes or coding.

Results

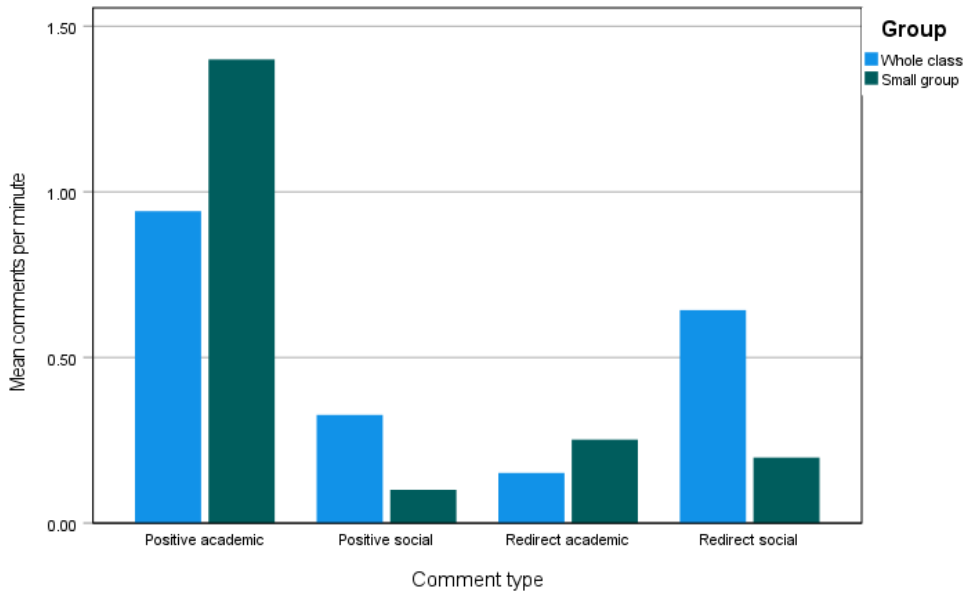
Descriptive Statistics

Mean frequency rates showed that within the whole class context, the most frequently occurring type of comment was positive academic (0.94 comments per minute) and the least

frequently occurring comments were redirective academic (0.15 comments per minute). Equivalent statistics for the small group showed that the most frequently occurring type of comment was positive academic (1.4 comments per minute) and the least frequently occurring comments were positive social and behavioural (0.1 comments per minute). A graph for the mean rates of comments per minute are shown in Figure 2.

Figure 2

Mean of Comments per Minute by Comment Type and Group



Accuracy of Teachers' Predictions. Descriptive statistics were used to consider whether teachers were able to anticipate the change in language usage between differently sized groups (hypothesis five). All data were entered into Excel version 365. Differences were calculated between predictions for small group use of language and whole class use of language. Mean differences showed that teachers correctly predicted that two types of comment would increase (positive academic comments prediction + 9.12%, actual + 26.92% and redirective academic comments prediction + 10.42%, actual + 6.67%) when moving to a small group context. They also correctly predicted that two types of comment would decrease (positive social comments

prediction -11.67%, actual -10.5% and redirective social comments prediction -7.92%, actual -23.25%) when moving to a small group context. These predictions were in line with the observed use of language data but strongly underestimated the magnitude of the change in positive academic comments and redirective social and behaviour comments.

Table 5

Predicted and Actual Differences in Proportion of Comments When Moving to a Small Group

Teacher	Predicted difference in positive academic comments	Actual difference in positive academic comments	Predicted difference in positive social comments	Actual difference in positive social comments	Predicted difference in redirective academic comments	Actual difference in redirective academic comments	Predicted difference in redirective social comments	Actual difference in redirective social comments
2	45	35	5	-10	-10	14	-40	-39
3	-10	15	10	-2	0	-4	0	-11
4	20	43	-20	-10	20	-1	-20	-32
5	-30	29	-15	-7	45	11	0	-33
6	5	14	-15	-2	20	7	-10	-19
7	5	65	-20	-27	25	5	-10	-43
8	5	2	0	5	0	1	-5	-8
9	5	10	0	-5	0	5	-5	-10
10	0	42	-10	-18	-5	7	15	-31
11	25	10	-15	-14	5	17	-15	-13
12	10	39	-35	-23	30	11	-5	-27
13	30	19	-25	-13	-5	7	0	-13
Mean	9.17	26.92	-11.67	-10.50	10.42	6.67	-7.92	-23.25

Descriptive statistics were also used to consider the accuracy with which teachers were able to reflect on their own use of language for each type of comment (hypothesis five). Differences were calculated between their predicted and observed use of language without consideration of direction to ensure that teachers did not cancel each other out. Total accuracy

scores showed that when teachers were asked to reflect on their whole class language, they were most accurate when considering their use of academic redirections and least accurate when reflecting on their use of social and behavioural redirections (where they underestimated the rate of change).

When asked to reflect on their small group language, they were most accurate when considering their use of social and behavioural redirections and least accurate when reflecting on their use of positive academic comments (where they underestimated the rate of change).

Teachers generally overestimated their use of

- positive social and behavioural comments in the whole class situation (number of teachers =9),
- positive social and behavioural comments in the small group situation (number of teachers =10),
- redirective academic comments in the whole class situation (number of teachers =11)

and underestimated their use of

- positive academic comments in the small group situation (number of teachers =10)
- redirective social and behavioural comments in the whole class (number of teachers =11)

Statistical Analysis

All data were entered into SPSS version 28 and were checked for the assumptions of an ANOVA to ensure it was possible to draw conclusions from the results. Box and whisker plots showed that two groups had potential outliers (whole class positive social and behavioural comments, and small group redirecting social and behavioural comments. In both cases, the outlier was higher than other scores, demonstrating that the participant concerned used a greater frequency of this type of comment than the other participants). Visual inspection of histograms for these two groups showed that it was one data point in each case but, because there are only 12 data points in the study, it is difficult to interpret or comment on more systemic patterns. However, analysis of z scores showed that these potential outliers were still within 99% of scores

in a normal distribution (Field, 2018) and therefore were kept within the analysis. Results were further analysed using the Shapiro-Wilk test for normal distribution and the Mauchly test of Sphericity. Two of eight data sets violated the Shapiro-Wilk test (those groups where z score checking had been necessary) but the ANOVA is considered robust to violations of normality (Blanca et al., 2017) so these remained part of the analysis. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi 2$ (5) = 1.410, p = .016, and therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε = .526). The Greenhouse-Geisser offers a more conservative correction than the Huynd-Feldt correction, thus reducing the likelihood of a type one error. (See Appendix J for all tests of assumption and ANOVA results.)

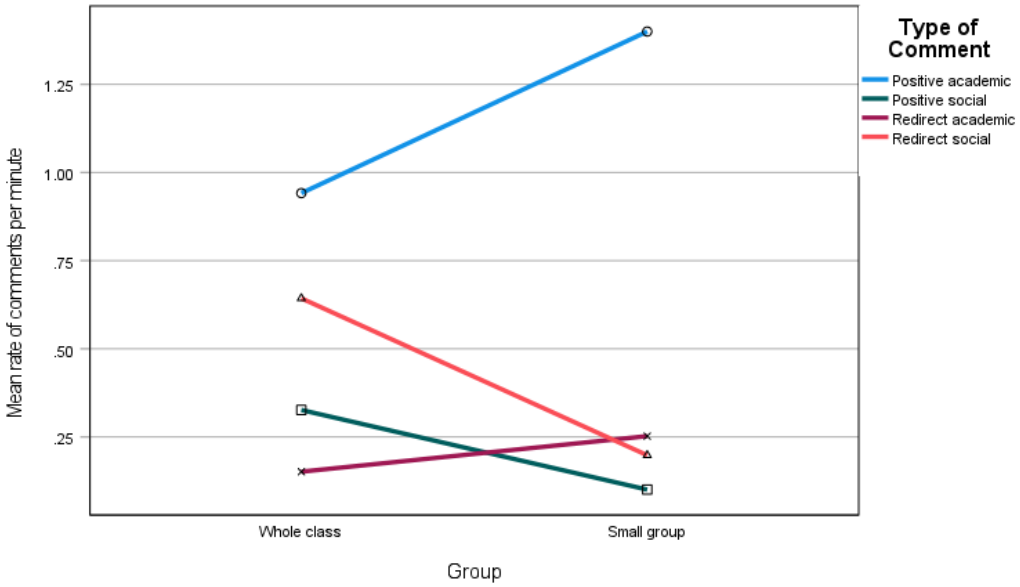
Mean of Comments per Minute by Comment Type and Group

A 2-way repeated measures ANOVA was performed to investigate whether there were significant interaction effects for the factors "group size" and "type of comment" on the rate of comments per minute (hypotheses one to four). The frequency-based "comments per minute" dependent variable enabled quantitative analysis across different lengths of observations (e.g., teachers' small groups were not all the same length). Simple main effects analysis found that when all four types of comment were considered together, there was no significant difference between the large and small group in terms of the rate of comments per minute (p = .331) but that there was a significant difference in the types of comments teachers were using regardless of group size (F (3,33) = 62.66, p < .001).

There was a statistically significant interaction between the effects of group size and type of comment (F (3, 33) = 31.12, p < .001). Whilst all data contribute to the significant model the greatest changes were the increase in positive academic comments and the decrease in redirective social and behavioural comments, as shown in Figure 3.

Figure 3

Graph of Type of Comments per Minute in Different Contexts



Effect size for the main effect of type of comment was $\eta_p^2 = .86$ and effect size for the interaction effect of type of comment and group size was $\eta_p^2 = .74$. Both of these demonstrate large effect sizes.

Four repeated measures t-tests were conducted to explore which of the four types of comment had changed significantly. (All assumptions for t-tests were met. See Appendix K for statistical analysis.)

There was a significant difference for each type of comment but not always in accordance with the direction predicted in the hypothesis. When teachers were working with a small group, there was a significant increase in the number of positive academic comments per minute, $t(11) = -5.47$, $p < .001$, $d = -1.58$, and in the number of redirective academic comments per minute $t(11) - 2.82$, $p = .008$, $d = -.81$. There was a significant decrease in the number of positive social and behavioural comments per minute $t(11) = 4.21$, $p < .001$, $d = -1.21$ and the number of redirective social and behavioural comments per minute $t(11) = 5.39$, $p < .001$, $d = -1.56$.

Two repeated measures t-tests were used to consider whether there was a significant difference across contexts when considering all positive comments regardless of reason or when

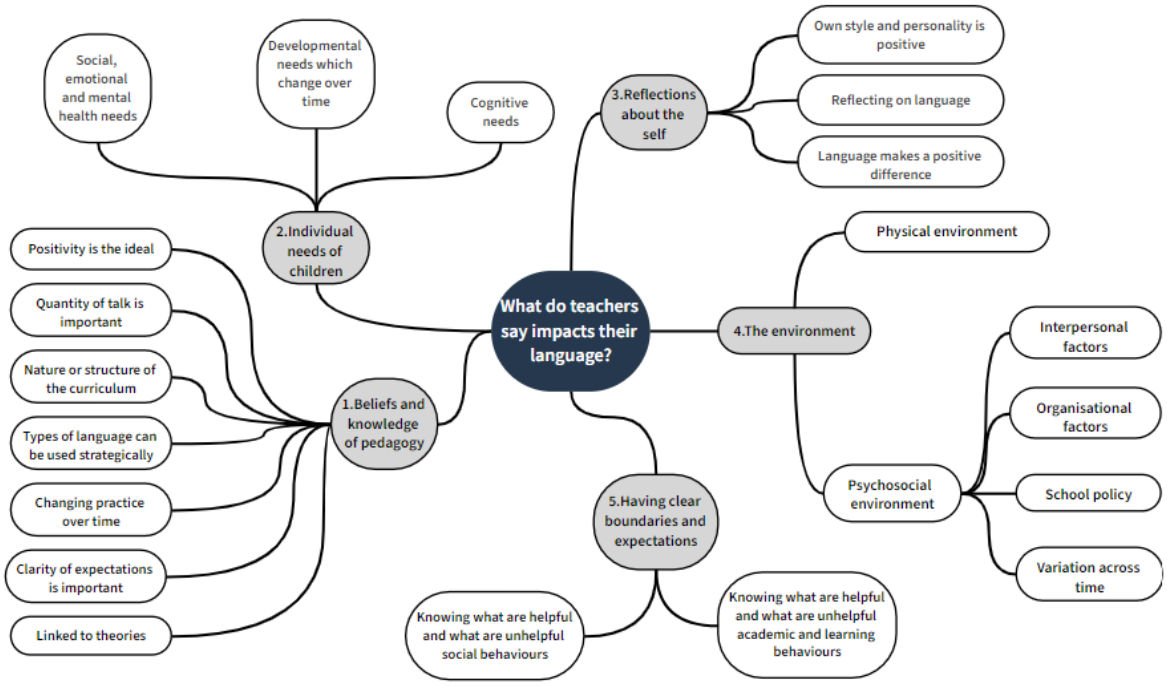
considering redirective comments as a whole. There was no significant difference in the number of positive comments per minute between whole class and small group contexts $t(23) = -1.35$, $p = .191$. There was, however, a significant difference in redirective comments when compared across contexts $t(23) = 2.39$, $p = .025$ although it should be noted that the redirective data violated the Shapiro-Wilk test of normality (see Appendix L for statistical analysis).

Thematic Analysis

Five analytical themes were generated to answer the question 'what do teachers say impacts their language?' (See Figure 4.) Each of these themes comprised a number of sub-themes, the most significant of which are discussed here. For details and examples of every theme, see Appendix M

Figure 4

Thematic Map



Theme 1: Beliefs and Knowledge of Pedagogy. This theme incorporated beliefs about method, style and delivery of the curriculum as well as psychological theories underpinning this practice. Teachers also reflected on changes over time both in their own practice and in school culture.

Positivity is the ideal. There was a strong sense from nine teachers that being positive is the ideal within the classroom; they expressed disappointment when they felt they had not met this standard. Teachers did not articulate a specific percentage of comments which they felt was the correct standard but they suggested that they would have liked to have used more positive comments overall. They explain this in terms of wanting to create a positive environment in the class. **"**I'd like to see a shift in those [the redirecting comments] just to make them a bit more balanced in the right direction." (Participant 10)

> Everybody goes in with high expectations of what they're going to achieve in terms of keeping the negative language to a minimum, and the positive language to the maximum. (Participant 2)

Two teachers also wondered out loud whether keeping language positive would result in fewer redirections being necessary or whether the reason redirective language was relatively high in the international study was a consequence of the low levels of praise.

Quantity of Talk is Important. Ten teachers voiced their awareness of the amount of talk being used in the classroom, regardless of the type or category. They reflected on whether their instructions were clear enough, not praising unnecessarily, noticing when they begin to "ramble" (participant 12), and personal goals of not talking too much (participant 9). One teacher spoke of using non-verbal communication to praise children and four teachers made direct reference to their choice of using non-verbal signals as a way to redirect behaviour such as a hand on the shoulder or standing and waiting. One teacher explained that he starts by giving a verbal cue to a child but then follows up with non-verbal cues if reminders are needed. Four teachers spoke about giving a "look" which they were happy the children understood. "I did have to give him a few reminders, um, sometimes even non-verbal, just that look. He knows, it's just about catching his eye at the right time" (participant 9). Teachers also articulated their choices around how and where to use language, especially when it was of a redirecting nature.

> Because sometimes it's, you know, it's not the right thing to interrupt the flow of teaching just to redirect a particular child's behaviour. Often, it's the best choice I think to just do it privately, quickly, you go down and then you come back. (Participant 7)

Proximity of the child to the teacher was also suggested as a non-verbal way to ensure a pupil was on task, less distracted, or nearby to enable plenty of encouragement.

Nature or Structure of the Curriculum. Ten teachers reported that the curriculum itself influenced the language they used in the classroom. Teachers had a clear understanding of whether the purpose of the lesson was to consolidate parts of the curriculum which had already been taught or whether this was new learning for the children. They referenced the importance of knowing when certain topics were new or tricky and when the children might need additional support, further reassurance or academic redirections. One teacher framed this to the children as 'unpicking.'

> But today because it was new language, … new maths signs to them, um, I think it was at their level enough but me supporting them and just saying 'have a go, don't worry, then we'll talk it through and unpick it.' (Participant 13)

Similarly, one teacher suggested that designing tasks which were in small steps offered opportunities for lots of praise. Other teachers articulated how certain types of task, for example recapping learning, new learning, or independent learning, impacted the type of language they felt was required to support children.

Types of Language can be Used Strategically. This sub-theme incorporated the opinions of nine teachers who all felt that language could be used strategically. They referenced redirective language (quick redirects, whole class redirects, finding something positive even if the other half of the comment is a redirection), discretion and tact (subtle comments, suggesting working together, pretending they do not understand so the child explains), and positive language (encouragement, addressing unwanted behaviours by praising others nearby, positive feedback for responses, public praise). This sub-theme also included the importance that teachers place on academic redirections such as "you've got to address those things to be able to move forward" (participant 10) and "you need to be able to correct them if they've done things wrong or um, move them down the right path" (participant 11), suggesting that teachers believe academic redirections are crucial for moving children forward in their learning. This sub-theme linked to the sub theme of relationships where teachers spoke about making redirections in ways which still supported relationships within the classroom.

Changing Practice Over Time. Within this sub-theme, teachers reflected on both positive and negative changes that had happened within education and society over time. Two teachers felt that children's behaviours were not as good as in previous years and therefore more social and behavioural redirection was necessary. This also linked to a comment within the psychosocial environment theme, in which a teacher said that children used to be more respectful than they are now. However, teachers also recognised positives over time, suggesting that teachers should not "rule by fear" (participant 13) as had been the case when they were at school, and another said there was a greater focus on helping a child understand where they went wrong, rather than it just being wrong as it had been in previous years.

Also in this theme were comments expressing teachers' willingness to have people "pop into your lesson saying 'ooh, I like the way you did this,' or 'have you thought about that'" (participant 10), to reflect, and to change their practice over time.

Theme 2: Individual Needs of Children. All 12 teachers made clear reference to adapting their language based on the individual needs of children. "So, it's just that isn't it, having that awareness of the children in your class, their needs, how they respond to you as a person" (participant 11). During analysis, these needs were categorised into cognitive needs, social, emotional and mental health needs (SEMH) or needs that were linked to developmental changes over time.

Developmental Needs Over Time. Five teachers articulated their thoughts about the age of the children impacting the types of language they used. They conveyed thoughts around children in Early Years or Year One needing a greater emphasis on emotions and supportive learning behaviours than older pupils. They also spoke about speaking out in class being harder when children are younger. Two teachers specifically referenced, with disappointment, the reduced amount of praise once children reach year six.

> I think with the positive comments about social skills and behaviour, I think that is sadly something that doesn't get… I do do a few like 'well done, thank you, you can get a tribe point' and trying to do positive reinforcement, um, but with the

> year sixes I think that sadly we don't do enough of that, um, praising those that are doing the right thing, just because it's seen as like, they're in year six, you know. I know when I'm down in lower school I do do a bit more of that.
>
> (Participant 11)

Another teacher commented that they "flip" (participant 13) disapproving comments into positive comments when working with younger children. One teacher felt that children became less resilient as they got older but was unsure why that might be the case.

Cognitive Needs. Eight teachers articulated how they were aware of children who might struggle with the academic content of a lesson and may need individual support or further explanations. Two teachers spoke about children who grasped concepts very quickly and then needed additional work. Targeted questioning was also discussed as a language strategy by teachers who felt they knew their children well and wanted to check the understanding of individual pupils. "That's sometimes why the no hands up, um, is really helpful, because as a teacher it's not a random picking something out, it's, I know who I want to ask certain questions to and why" (participant 7). It was also considered a helpful way of sharing good ideas from which the rest of the class could learn, for engaging children, and helping individuals to concentrate.

Social, Emotional and Mental Health Needs. Similarly, all 12 teachers reported that the social, emotional and mental health needs of individual pupils in their class impacted their choice and use of language. Six teachers discussed their use of encouragement or praise to boost a child's confidence levels. "I just feel like my job for her is just to cotton onto anything that she's done really well, and just fill her up with praise and boost her confidence" (participant 2).

Teachers were also aware of which children might need extra encouragement. "It's just frustrating for me because I know how much he can do. Um, so yeah, I was kind of geeing him on a little bit" (participant 9).

One teacher spoke about following a child's lead as to whether to push an academic agenda or be more accommodating of their emotional needs on a given day. For example,

participant six was aware of the pupil's background and past experiences. "Because she started off writing something, and then she stopped and got distracted and wanted to talk about [sensitive subject], which I'm not gonna stop her talking about…, but she didn't do what I wanted her to do" (participant 6).

As with cognitive needs, one teacher spoke about how they knew in advance that a pupil had a good answer and therefore encouraged the child to share it with the class. This was to boost the pupil's self-esteem and give the teacher the opportunity to publicly praise him.

Theme 3: Reflections About the Self. This theme incorporated comments from 11 teachers who spoke about their reflections regarding themselves. These teachers commented on their approach within the classroom and their beliefs about positivity being part of their self-identity. This theme also included comments on the ease, difficulty or usefulness of reflecting on language use in the classroom.

Own Style and Personality is Positive. Six teachers felt that the personality of the teacher impacted the language they used. They commented on their own nature, style and preferences. "I think … as a person … I prefer to be more positive than negative" (participant 8).

One teacher also commented that they maintained their positive style and approach even if children's comments were a little off topic because they "give a lot of positive praise. That's often just how I am, I like them to feel like they are in a safe space to learn and just enjoy the lesson" (participant 11). Another teacher commented that they made a particular effort to always be positive because they remembered experiences from their own time at school which were not positive. They recognised this motivation to ensure that they always took a positive approach and sought to address all situations in a positive way, even if behaviour or learning was not as they wanted it to be in their class.

Reflecting on Language. Seven teachers considered their awareness of the language they used and reflected on whether it varied in different circumstances. One teacher felt that

redirecting language is very natural teacher language and not always in conscious awareness whereas being positive is more memorable.

> I think you're less aware of how much you probably do the 'oh make sure you're sitting properly' or 'make sure you're doing your um letters the right way, doing your finger spaces' cause I think that's all kind of - it's - it comes as more of an automatic thing. (Participant 3)

Other teachers agreed that their use of language feels automatic and beyond conscious awareness, and that reflecting is "tricky" (participant 13). Despite this potential difficulty, teachers valued the reflective opportunity and were keen to consider their use of language as part of their professional development going forward.

> And I've found it really helpful, …you don't get that opportunity to have that data presented to you. It's often our own personal judgement as teachers, you know, which can clearly vary quite a lot with the reality, um, and I'd rather know what the actual reality looks like so I can go and take that and reflect on that and decide whether there's any other kind of choices that I need to make next time. (Participant 7)

Language Makes a Positive Difference. Five teachers recounted specific situations in which their language had made a positive difference. "She also sort of struggles, but she was grasping it, but she also did need a bit of help so, it felt like I was actually helping her and teaching her" (participant 4). They were able to pinpoint the language and actions they had used to make a difference for a child, whether that was praising a pupil's work to the class as an exemplar, quietly praising a child in a small group, or commenting on a 'brilliant idea' before writing it on the board for all to see. Teachers made their judgements around the impact of their language through observation of the pupil then trying really hard, seeing small steps of understanding even if the child did not fully grasp the concept, and in one case, a boy silently turned his traffic light card to green to indicate he was feeling more positive about the learning.

Theme 4: The Environment. Although the physical environment and the importance of a focused and quiet classroom was articulated by two teachers, the emphasis in this theme was on the psychosocial environment with all teachers discussing this in some way.

Psychosocial Environment: Interpersonal Factors. Three teachers spoke about the importance of relationships within the classroom. Teachers wanted to support relationships between peers, to build relationships with pupils in order to break down barriers to learning, and to ensure that pupils felt valued by their teacher.

> If teachers aren't saying to children 'oh well done, great effort, I'm really proud of you, I can see you're trying really hard, I know it's tricky but you know, you're trying your best, thank you,' all of that, it's about relationships, isn't it? Teaching is all about relationships. (Participant 13)

One teacher also linked their comment about relationships back to their expectations of behaviour and how to strategically use language as discussed above.

> Because it's not about humiliating children or, you know, it's just about having that quiet word and actually, it's all going back to relationships - if those children want to work for you and get that positive [praise], then that little quiet word is enough to make them turn it around, because they actually want you to be proud of them and want for you to say well done. (Participant 13)

The sub-theme of relationships tied closely to the sub-theme of using language strategically in the beliefs and knowledge theme. Teachers were very aware of using language strategically to address behaviour concerns positively or suggest tackling difficult tasks together; they were also aware of merging a positive comment and negative comment together, for example telling a child to turn around so they could see their lovely work. This provided further evidence of their desire to build positive relationships and environments within their class.

Psychosocial Environment: Organisational Factors. Eight teachers spoke of how their choices regarding the organisation of the learning environment (using small groups) impacted

their language. “Because it’s a smaller group there’s… I felt there was hardly any redirection in terms of behaviour” (participant 3). Small groups were felt by participants to enable more positivity, more focus on the learning tasks, and greater confidence building in the children.

Psychosocial Environment: School Policy. Four teachers referred to whole school policies and expectations when considering their use of language. These teachers felt that their schools encouraged a positive approach as well as high expectations.

> I do feel that I do do a lot of having to pick people up … if they’re not listening, or … they’re not, you know, behaving in the way that our school expectations expect, and this is very much led by school culture and expectations, um. As well as my own, but I’m very aware of the expectations in school, I guess, and what we what we [sic] want from the children. (Participant 11)

Psychosocial Environment: Variation Across Time. This theme incorporated reflections from ten teachers about changes across decades, changes since the COVID-19 pandemic, and changes that occur each year such as the long summer holiday. When considering changes across decades, one teacher felt that expectations within the classroom had not changed, but expectations in the home environment had. This led to more redirection of social skills and behaviour than in previous years. Another teacher suggested that the balance of behaviour focus, and academic focus had shifted.

> I am aware having taught through both periods [referring to the statistics from previous research in 2000 and 2019] that in 2000 I would’ve said I would’ve been doing less um looking at the behaviour and more being able to, uh, look at the academic side than I do now. So, there are behavioural changes. (Participant 6)

Another teacher questioned whether there had actually been a decline in behaviour or whether teachers were now more aware of the importance of children being ready to learn before they can focus on academic skills. One teacher specifically referenced the COVID-19 pandemic as having had an impact on children’s social skills such as sharing.

The time of year impacting teacher language and teacher focus was frequently mentioned, with teachers suggesting that early on in the academic year was a time to focus on routines and classroom expectations.

> I've found that it is that first half of the term, where obviously the learning and the task is really important, but actually a lot of my input is about building in those routines and those behaviours which then will then pay off throughout the year. (Participant 7)

As well as the focus of language changing across the year, one teacher made reference to children learning common non-verbal cues and thus fewer verbal redirections being necessary. "Whereas other children, throughout the year have had less and less verbal redirections and reminders because they learned the non-verbal cues" (participant 5). This links to the sub-theme around quantity of talk, as the teacher was choosing to replace the verbal reminders with non-verbal cues with the majority of the class.

Theme 5: Having Clear Boundaries and Expectations. This theme encompassed teachers having a clear idea of the behaviours that they would praise and encourage, along with the behaviours they did not wish to see and would strategically try to reduce, whether through language or other means.

Summary of qualitative and quantitative results

The congruent mixed method design meant that whilst quantitative data were gathered prior to qualitative data for individual teachers, the study as a whole did not place precedence on one or the other. This meant participants reflected on their own beliefs and perceptions in order to offer a deeper level of insight into the observational data of their own classroom practice. The study as a whole then considered how these two types of data interlinked.

The first research question in this study focused on the content of teacher language and how it was distributed across four categories. The second research question considered whether teacher language changed depending on the context (whether whole class or small group). This

study found that, in a whole class situation, the most frequent comment was an academic approval, followed by disapproval of social behaviours. This linked to teachers' qualitative explanations when they discussed their particular beliefs (e.g., feeling like they were a positive person and knowing individual pupils well). Analysis of the data also showed that there is a significant interaction between type of language used and the context. The third research question explored whether teacher perceptions and recollections of their language use was in line with observational data. Teachers were able to accurately predict direction of change for some comments but also overestimated and underestimated frequencies. A deeper level of insight into these statistics was engendered when teachers spoke about positivity being an ideal ethos and about how they valued opportunities to redirect pupils' learning. The fourth research question offered greater understanding of a wide variety of factors which teachers say impact their language in the classroom.

Discussion

The current study aimed to explore the content of teacher language and how this changes across contexts, the extent to which teachers can reflect on the language they use, and the reasons teachers give for the choices they make around language.

Research Questions One and Two

The current study has added to the understanding of how teacher language changes in a small group situation. Statistical testing found that context (whole class or small group) significantly interacted with the types of comment being made by teachers. When teachers moved into a small group context, positive academic comments significantly increased (hypothesis one accepted) and redirective social and behavioural comments significantly decreased (hypothesis four accepted). This suggests that teachers adjust their language rates according to the environment in which they are teaching. Whilst teachers were able to recognise (post observation) that their language rates had changed in the different contexts, this had not been a conscious decision for them during the lesson. It appears that teachers adapt their

language according to their environment, even though they may not be fully conscious of doing so.

In a similar but not identical study, Floress et al. (2021) observed teachers in grades one to four in Illinois, USA, and calculated mean rates of praise per hour and reprimands per hour in whole class contexts and small groups. Analysis using t-tests found no significant difference in rates of positive comments when changing between a small group and whole class. Neither was there a significant difference in reprimands between contexts. However, in the current study, a significant difference was found in the number of redirective comments per minute (combining academic and social comments) when teachers changed contexts. Furthermore, analysis in this study paid particular attention to the reason for the positive or negative comment (i.e., whether it was regarding academic or social behaviours) rather than grouping them into 'general' or 'specific' categories and found an interaction. It may be that the finer level of coding and analysis in this study allows greater insight into teachers' use of language.

Classroom environment and positivity make an important difference to pupil behaviour and outcomes (Alber & Heward, 2000; Henderlong & Lepper, 2002) so when schools and teachers are making decisions about lesson planning, it is important to consider that teachers' natural rates of positive academic language increase in small groups and redirective social and behavioural comments decrease. This means small groups have a definite role to play in increasing opportunities for positivity. Similarly, as academic praise and academic redirections significantly increase in small group contexts there is a helpful function for them when considering academic goals in the classroom.

Research Question Three

In this study, teachers' predictions were consistent with the direction of change for all four types of comment, suggesting that teachers have some awareness of how their language changes across contexts. Indeed Floress et al. (2022) also found that teachers were able to consistently match their predictions of praise and reprimands with data from 20 minute observations. Although some individuals had difficulty, and Kim & Stormont (2016) found

inaccuracies with reflections of two-hour periods in Early Years settings, there appears to be potential for post lesson reflections to offer helpful and accurate starting points for discussion and professional development. Being able to reflect also enables teachers to foster the practices and skills required for self-directed growth (Zeichner & Liston, 1987).

Teachers overestimated their use of positive social and behavioural comments in both the whole class and small group situations and underestimated their use of redirective social and behavioural comments in the whole class. When seeking to understand this, a theme in the teacher interviews was the idea of positivity being the 'ideal.' Floress, Beaudoin and Bernas (2022) suggest that teachers know they should be using positive comments and therefore report that they are. If this is the case, impression management (deliberate attempts to make oneself appear better or more desirable) may have played a role here. On the other hand, teachers knew that the researcher had collected observational data and therefore manipulation of their own data would be ineffective. Therefore, self-deceptive enhancement may be more likely in which participants offer honestly held views but these views are inaccurate or unrealistic (John & Soto, 2007).

In this study, teachers also overestimated the number of redirective academic comments in the whole class context. In other words, teachers did not make as many academic corrections as they thought they had. In their interviews, teachers valued academic redirections and felt they were helpful in moving children forward in their learning. This discrepancy may be a further example of self-deception enhancement but may also link to explanations offered about targeting questions. In their interviews, teachers spoke about using well-calibrated questioning to enable pupils to demonstrate competence in front of the class. Sometimes teachers were aware in advance that a child knew the correct answer and utilised this for reasons such as building self-esteem. By utilising finely tuned questions, teachers reduced the possibility of incorrect answers and thus the need for academic redirections in the whole class context.

Research Question Four

Teachers in this study suggested that the age of pupils impacted their language and reported that higher rates of praise were more common in younger age groups. This trend of

decline in rates of praise as children get older is also suggested in international research (Drake & Nelson, 2021; Floress et al., 2018; McLennan et al., 2020; Reddy et al., 2013; Sulla et al., 2019). Although teachers in this study accepted it was the case in their classrooms, they expressed concern. One teacher particularly commented that she realised she used more praise when teaching younger children but she regretted the fact that older pupils received lower rates of praise. An explanation from one teacher was that less praise was given to older pupils because older children already know when they are doing the right thing. However, teachers' disappointment suggests they still believe that praise is important for all age groups and that they were previously less aware of their language distribution in the classroom. Greater knowledge about their language distribution did not match their internal beliefs either around positivity being the ideal type of language, or being a positive person, or that they focus on the needs of individuals (who may need more praise than others regardless of age).

Reduced praise in the classroom for older pupils, whether a conscious or unconscious choice, may be unhelpful because placing greater focus on correct and appropriate behaviours can create a more positive and enjoyable learning environment (Caldarella et al., 2020; Drake & Nelson, 2021; Spilt et al., 2016). This in turn can help to avoid difficult situations in which inappropriate behaviour is reinforced through attention, especially for 'at risk' pupils (Downs et al., 2019).

In this study, many teachers spoke about their positive attitude and drive for professional development. Teachers in their interviews discussed how willing and open they were to engage in continual professional development. An important explanation from all teachers regarding their choice of language was consideration of the individual needs of children; a wide variety of children's needs were voiced and teachers were very aware of selecting differing approaches for different children. This is supported by Sabey et al. (2019) who argue for the importance of adapting language based on the needs of individuals. In their work discussing the concept of optimal positive to negative ratios, they explain that different ratios (from 1:1 to 10:1) have been suggested in the past 25 years, yet there does not appear to be unanimous agreement as to what

an ideal ratio might be. Instead, Sabey et al. (2019) advise that teachers monitor their interactions to ensure they are achieving the desired intention. Teachers "may also find it helpful to vary their PN [positive negative] ratio based on the needs and performance of individual students" (Sabey et al., p. 162). This would suggest, for example, that if the work is new or challenging for a pupil, teachers consider higher rates of additional positive praise to support resilience or independence; this is indeed what teachers in this study explained they were consciously doing.

Sabey et al. argue that it is reductive to focus on a particular ratio, and that instead teachers should focus on a desired intention, which may be supported by a praise ratio that is prompted from the circumstances they create. This avoids focusing too much on teacher behaviour and language in the moment (teacher performance) and allows a renewed focus on creating the circumstances for improved pupil performance.

This study considered the frequencies of different comments rather than converting the data into ratios, which would have been a different way to represent and understand the same data. Regardless of the format, both approaches speak to the ethos of the classroom environment experienced by children and teachers. Ratios may have the advantage of being more easily tracked, monitored and adapted in the moment for teachers, however, this data-driven approach can lack a focus on individual circumstances (as discussed by Sabey et al., (2019)). The comments per minute approach in this study enabled an in-depth analysis of the importance of context and circumstances for influencing teacher rates of language in the classroom. This type of analysis would not be possible for teachers 'in the moment.'

'Withitness,' Smoothness and Momentum. Kounin (1970) argues that effective teachers demonstrate "withitness " which is when a teacher communicates to the class through their behaviour that "she knows what the children are doing, or has the proverbial 'eyes in back of her head'" Kounin (1970, p. 81). Effective teachers can deal with two things at once by "overlapping" and can respond quickly to stop ripples of disruption. Kounin also suggests that effective teachers notice (and show they have noticed) inappropriate behaviours but then choose to continue with

their current action. This ensures smoothness and momentum within the lesson which in turn correlates with positive pupil behaviours (Kounin, 1970).

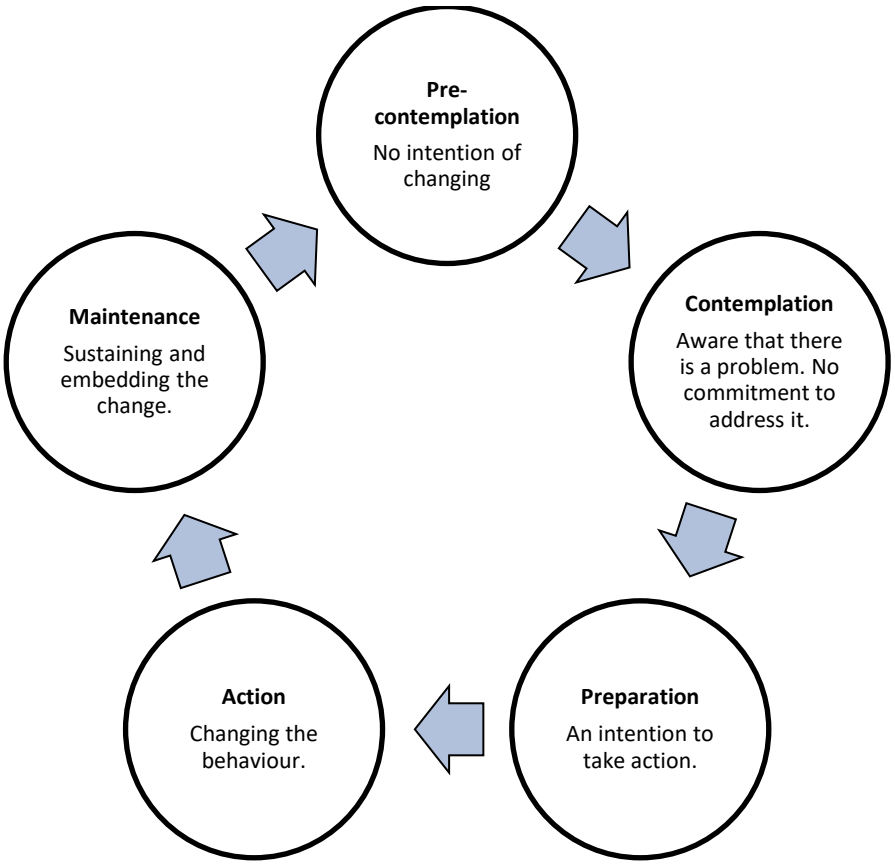
One sub-theme in this study was titled 'quantity of talk is important.' The teacher who spoke about choosing not to "interrupt the flow of teaching just to redirect a particular child's behaviour" (participant 7) was unknowingly articulating Kounin's argument about ensuring momentum within the lesson. Another point in this sub-theme was teachers feeling that verbal communication was only part of their available toolbox. They believed that reducing the amount of teacher-talk in the classroom was an important goal, so they would choose to utilise non-verbal communication if possible and only to comment if necessary. This non-verbal communication made it clear to pupils that their behaviour had been noticed and was unacceptable (demonstrating Kounin's withitness) but did not interrupt the smoothness or momentum of the lesson. It also enabled the teacher to stop inappropriate behaviours before they spread or increased in seriousness. Using non-verbal communication is in line with work by Jones et al. (2007) who argue that teachers have great power in their use of non-verbal communication (including eye contact, proximity and gesture) and can utilise these subtle, interpersonal skills to support effective discipline in the classroom.

Implications for practice

In this section, implications for practice will be linked to the stages of change model (Prochaska & Di Clemente, 1982) in order to offer, not only implications from the study, but a reasoned order in which they may be helpfully implemented. This model offers five distinct stages in any change of behaviour (see Figure 5) of which four are pertinent in this research and will be described. Furthermore, throughout this section, a clear checklist will be developed for use in schools (see Appendix N).

Figure 5

The Stages of Change Model



The Pre-Contemplation Stage: Not Yet Recognising That There is a Problem

When considering the application of these findings to professional development for teachers, the results in this study raise possible difficulties if teachers are simply advised in training sessions to 'use more praise' in the classroom. Teachers in this study were already using more praise than redirective comments and were aware of this. In order for any professional development or change in teachers' behaviours to occur, teachers must first recognise that there may be a need for this. The first stage in the stages of change model (Prochaska & Di Clemente, 1982) is known as 'precontemplative.' During this stage, people are unaware that there may be a need for change and therefore do not move on any further through the model towards a change in behaviour. Teachers who believe that they are already using lots of praise in the classroom may not feel the need to move to the contemplation stage.

When considering teachers' over and underestimation of different types of language, self-deceptive enhancement is one potential reason. If this indeed is the case then teachers would still be in the precontemplative stage of change (Prochaska & Di Clemente, 1982). They would be

unaware that there is a need for them to make a change and therefore not believe that any form of 'language in the classroom' training was necessary or helpful for them personally. Consequently, any training would need to start with consciousness raising and even considering any potential defensiveness towards change.

One concern raised from these findings of inaccuracy in teachers' recollections is that teachers did not appear to be fully aware of how levels of language change according to context. Therefore, it would be helpful to ask teachers to gather information, observe peers or reflect on their practice and the types of language they use for different purposes before considering any weighing up of information as part of the contemplative stage. Indeed, there is a role for peer support when reflecting on language in the classroom. The importance of observing others and being observed in a supportive and developmental way should not be underestimated because the evidence here has demonstrated how complex it can be to recall the many interactions and comments made during a lesson and the potential limits of teachers' reflections on themselves. (Teachers were able to predict which types of comment would increase and which would decrease. However, they were subject to an underestimation of change bias for positive academic comments (which increased more than they expected) and for social and behavioural redirective comments (which decreased more than they expected.)) Therefore, teachers may need more support to reflect on their lessons and move on from the pre-contemplative stage. Then, when training does take place which asks teachers to consider changes as part of their approach to helping pupils learn, teachers would be better placed to move forward.

The Contemplation Stage: Acknowledging That There is a Problem

Findings also demonstrated that there were sometimes disparities between teachers' beliefs and hopes about best practice and their actual use of language (e.g., believing that older children still need high levels of praise). Therefore, gathering observation data, especially in upper primary year groups can bring language use more fully into the conscious awareness of teachers and offers the opportunity to highlight any discrepancies between teacher hopes for positivity and actual behaviours. Once aware and seeing a need for change (i.e., in the

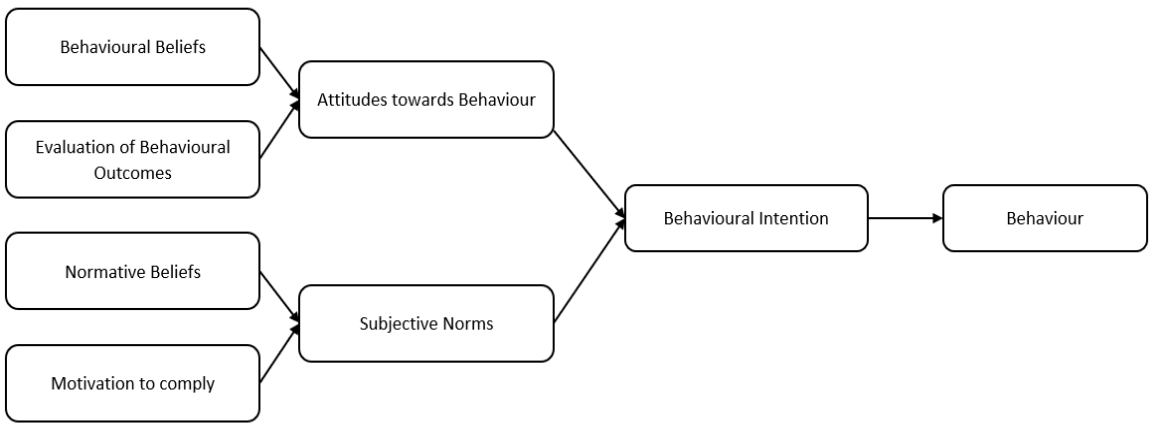
contemplation stage), teachers are better placed to set goals, utilise professional development and prepare to change as discussed in the stages of change model (Prochaska & Di Clemente, 1982).

Preparation Stage: Preparing to Make a Change

Awareness of quantitative observational data may support teachers through the pre-contemplative stage, into the contemplative stage (where they are aware that a change may need to be made). However, it is the third stage (preparation) in which a teacher must develop the intention to make a change. Teachers spoke of a ~~this~~ positive attitude and drive for professional development This links to ~~is~~ the third stage of the stages of change model (Prochaska & Di Clemente, 1982) and is a crucial underpinning of any action. The preparation stage ties closely to the idea of behavioural intention as described in the theory of reasoned action (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975) (see Figure 6). This model suggests that behaviours are driven by attitudes (personal) and subjective norms (social influence), which feed into a person's intentions, which then result in the observed behaviour.' This study generated some key themes which illuminate personal attitudes and influences which drive behaviours. Having a strong sense of identity around being a positive person and having a positive style formed a sub-theme within 'reflections about the self'.

Figure 6

Theory of Reasoned Action



Teachers' behavioural beliefs (e.g., that their choice of language makes an important difference to pupils, "it can only be something small that you say, but it does have quite a big impact on them" (participant 9)) lead to an attitude or disposition towards using different types of language. Teachers' normative beliefs (that their headteacher or school policy or pupils for example, wish them to use certain types of language, "it's something that we try to focus on a lot in our teaching … positive sort of environment and atmosphere" (participant 10)) and motivation to comply (how much they want to comply with those people, "this is very much led by school culture and expectations, um. As well as my own" (participant 11)) lead to a subjective norm which takes into account for all the different opinions. Together, attitude and the subjective norms lead to an intention to perform (or not) a certain behaviour, in this case, using certain types of language, or language in a particular way. This has implications for any professional development opportunities which consider language use in the classroom. If the ultimate aim of some training is to finely tune language usage, it would not be helpful to immediately focus on the behaviour at the end of the model. Instead, careful account must be taken of teachers' attitudes, motivation, beliefs around subjective norms and beliefs about behaviours and their outcomes. Without these considerations, intention and behaviour are unlikely to shift. Relating this back to the teachers in this study, there were clear references to having a positive personality and style (attitude) as well as recognition that positivity was in line with their school policy and their colleagues (subjective norms) with which they wanted to comply. This suggests that in order for training to be effective in promoting changes in language use, it should not just focus on behavioural change, and that it should also consider the beliefs and norms that exist for individuals and in the school context.

Action Stage: Take Steps to Change Behaviour

Once the initial three stages are complete, teachers are well placed to make professional development changes. The findings of this study suggest three key areas for change.

Small Groups Make a Difference. The implication from this study is not that small groups should be used at all times, simply that they offer opportunities that should not be ignored when

aiming to influence rates of language in the classroom or when considering wider pedagogical recommendations (e.g., the benefits of whole class reading). Small groups can and do offer variety within the school day, helping to reduce or avoid satiation of any one type of teaching methodology, which is when repetition of an activity results in increasing dislike and becomes "less and less positive" (Kounin, 1970, p. 126). Teachers should therefore be alert to the benefits of small groups and when and why they can be helpful; careful consideration of the learning context is a valuable and important part of lesson design and planning. Therefore, when teachers are planning a lesson, or considering their priorities for certain classes or children, the context (small group or whole class) needs to be carefully considered and small groups recognised as a helpful environmental strategy when helping children to learn.

Environment and Individual Needs (not Ratios). This study demonstrated that teachers' language changed as they moved from whole class to small group teaching, in that they increased their positive academic comments and their academic redirective comments. When academic competence is the objective in a lesson, small group opportunities appear to offer an environment in which teacher language will change to include more focus on academic content. An implication of this current research is that, if they are aiming to influence their rates of language use in the classroom, (for example increasing praise and approval rates to increase pupil time on-task (Swinson & Harrop, 2010) which in turn may be linked to better pupil outcomes (Fisher et al., 2015) or to promote a more friendly and secure environment (Spilt et al., 2016)), teachers could better focus on creating the context which prompts their desired rates, rather than focusing on aiming for a particular praise ratio.

'Withitness,' Smoothness and Momentum (Kounin, 1970). As discussed above, teachers valued and recognised the importance of non-verbal communication and keeping a lesson flowing. Whilst teachers in this study did not articulate the term 'minimal sufficiency', their aim for no unnecessary language is consistent with ~~this~~ the minimal sufficiency principle ~~too~~, in which subtle or minimal social guidance is given but which eventually leads a child to adapt their understanding of social norms and values (Lepper, 1985). Taking action in terms of developing

skills to enhance smoothness and momentum within the classroom can therefore be a helpful component of professional development focused on language use.

A Comprehensive Checklist

In order to support the implementation of this research, a comprehensive checklist was developed (see Appendix N). The checklist is designed to be used through supportive peer opportunities for the purpose of individuals' professional development, and in advance of any specific training. If whole schools utilise the checklist, it will provide an opportunity for school leadership teams to identify areas for school development plans, but importantly with the knowledge that teachers had discovered these areas for development themselves.

The checklist is broken down into four steps, reflecting the four stages from the stages of change model discussed in this research and shares the findings and wider literature links in an accessible way. Step 1 invites teachers to reflect on their own language within the classroom and discuss this with a peer. Step 2 suggests teachers consider if there are already any areas in which they feel they might like to improve. Step 3 asks teachers to reflect on their own attitudes and motivations. Finally, step 4 invites teachers to consider what they are already doing to enable learning and to create a positive environment, why it is important, and offers suggestions from wider research and this study for further steps to support this: (a) use of small groups, (b) ensure adapted language for individuals (Sabey et al., 2019) and (c) utilise non-verbal communication to support momentum, smoothness and 'withitness' (Kounin, 1970). All of these are in line with teachers' own hopes and beliefs about their use of language.

Strengths and Limitations

Strengths. A strength of this study was the mixed methods design, taking understanding beyond the statistics and illuminating quantitative results with teacher perspectives and explanations. The semi-structured interviews elicited detailed and thoughtful accounts from teachers with the topic guide enabling both control and flexibility. The inductive, reflexive approach to thematic analysis meant a deeper understanding of teachers' experiential views,

whilst the study itself retained a critical realist outlook, recognising that those views are shaped by social factors (Coyle, 2021).

Whilst the sample size was small, it offered a range of school communities, a wide span in terms of teachers' years of experience and a cross section of pupil year groups. In this study, 75% of participants were female which exactly matches the school teacher workforce statistics (Department for Education, 2023).

This study is also able to offer clear implications which argue that it is important to focus on understanding teachers' perspectives and the accuracy of their reflections before considering professional development. This is a crucial first step. Furthermore, this study offers more detail and nuance around 'praise training,' offering details on how teachers' language may adapt in different contexts and how different types of language can shift in different ways.

Limitations. There were a number of imitations to this study which mean that generalisations should be made with caution. One of these was the small sample size of both schools (n = 5) and teachers (n = 12). Whilst the sample did cross LA borders (two LAs), all schools were within forty miles of each other. Some teachers worked at the same school as each other and were therefore subject to the same school policies which may have led to some thematic codes being similar for contextual reasons. The small sample size also meant that despite some reoccurring themes with the thematic analysis, it is not possible to be confident of saturation.

Further limitations arise from the individual nature of the research. Firstly, only one researcher conducted the classroom observations. The researcher used a pilot study and table of example decisions (see Appendix F) to support reliability in observation decisions, but it was not possible to compare decisions made when completing the OPTIC. Similarly, individual teachers' understanding of definitions of the types of language may have varied. As discussed with regard to the pilot study, terminology (i.e., whether language was referred to as negative or as redirective) can have different emotive and semantic implications. Whilst every effort was made to explain terminology clearly to ensure consistency, it could be complex. For example, instructing an Early Years pupil how to sound out a word was not included (it was not based on a past event),

but if the 'sounding out' needed to be repeated by the teacher because the child used incorrect graphemes, it was included as an academic redirection. Despite efforts to clarify the terminology used, slight differences in understanding of definitions may have impacted the ways teachers categorised their use of language.

When considering the thematic analysis, one researcher completed the thematic coding which meant there were fewer opportunities to reflect on the meaning, expansion, and application of codes. Doucet & Mauthner (2012) further argue that the data analysis stage is the time at which the "power and privilege of the researcher are particularly pronounced" (Doucet & Mauthner, 2012, p. 127). To try to overcome these limitations, careful discussion with an experienced supervisor took place.

This study did not prescribe to schools which lesson would be observed nor the timing within the day. This meant that observations sometimes included transition times or settling times or, in the case of Early Years, free flow times. Therefore, some teacher language will have been based on pragmatic and logistical concerns and context may have necessitated different types of teacher language.

The research questions did not narrow the focus down in order to investigate one specific factor which may influence teacher expectations and language in the classroom. Reviews of the literature in this area (e.g., Wang et al. (2018)) showed that gender is a potential influencing factor in teacher expectations yet the evidence is currently inconclusive. Some studies suggest teachers have higher expectations for girls in literacy and boys in mathematics. Other studies suggest teachers have higher expectations for girls overall whilst different research has not found any impact of gender bias in the expectations of teachers. A wide variety of factors could be influencing teacher expectations and impacting language in the classroom, and consideration of any one of these (e.g., gender bias) would have added further insights and nuance to this research. Further work should be done in this area.

Fleming et al. (2016) suggest that humans are able to make judgements about their own performances. These evaluations are frequently gathered through asking participants to make

assessments of their confidence in the accuracy of their recall (Koriat, 2007) and these evaluations are often known as confidence judgements. Once gathered, they can be calibrated against data from the participants' actual performance (Dinsmore & Parkinson, 2013). Gathering such confidence judgements (e.g., using a Likert scale) could have been an additional element within this study which would have offered the opportunity to analyse and consider differences between teachers, and between teachers and their evaluations of themselves. Furthermore, Bandura's (1986) reciprocal determinism model has been used to explain confidence judgements (Dinsmore & Parkinson, 2013) with the suggestion that three different factors may be implicated and interacting when people make a confidence judgement: the person (e.g., prior knowledge), the task, and the environment. The current study sought to explore the factors impacting language in the classroom. Therefore, whether any of the factors in the reciprocal determinism model impacted any participant's confidence judgement could have been a relevant, linked, and potential next step within the study. The quantitative data available through calibration analysis alongside reflections on how the confidence judgements were reached could have added a further dimension and deeper understanding of the topic and further enhanced the mixed methods approach. This could be an appropriate and helpful next step for future research.

Finally, this study focused on verbal communication within the classroom and observation protocols were designed around this. However, non-verbal communication was raised by teachers in their interviews as an important part of their teaching toolbox as well as timing and quantity of language such as choosing when not to speak. The quantitative data did not capture any of these moments. Similarly, para-verbal communication was not monitored or captured during the observation with the caveat that it was occasionally used to help with researcher decision making (e.g., when a child's name was called out and the tone of voice made it clear that it was a redirection). All of these elements of communication are part of the emotional life of the classroom yet only the verbal features were captured.

Future Research

This study has provoked further questions about the accuracy of teachers' knowledge regarding their use of language in the classroom. Future research should consider investigating which factors predict accuracy in this area and how these can be effectively harnessed to support teachers in their reflections. These factors may include 'within teacher' reasons such as years of experience or beliefs about themselves. They may also include 'environmental' reasons such as group size or whether the lesson included transition times.

An unanswered question arising from this study is the difference in results when comparing it with the work of Floress, Zoder-Martell, et al. (2021). Both studies compared language rates across small group and whole class contexts yet found differing results. Further research would enable greater investigation into the nuance of differing language types and categories, and how these might change in different contexts.

An update to the Beaman & Wheldall (2000) review and analysis which showed the trend of increasing positivity over time would be welcome in order to consider whether this trend has continued in the 21st century. In the years since the review, there have been systemic changes to the UK education system. These changes include the academisation of schools, described as a slide towards authoritarianism (Reay, 2022) or an opportunity for greater autonomy, (Department For Education, 2010) and the new national curriculum in 2014 which was designed to match that used by the world's most successful education systems (Gove, 2013). Furthermore, the COVID-19 pandemic forced schools to restructure teaching and learning systems in unprecedented ways and this too may have impacted trends within the classroom.

Appendix A Systematic literature review search terms

TX (boy* or male*) OR (girl* or female*) OR (gender or sex)

AND

(“teacher* expectation*” OR “teacher* expectanc*”) OR ((“teacher* judg*ment*” AND (achievement* OR performance* OR outcome* OR abilit* OR attainment*))) OR ((“teacher* percept*” AND (achievement* OR performance* OR outcome* OR abilit* OR attainment*))) OR ((“teacher* belief*” AND (achievement* OR performance* OR outcome* OR abilit* OR attainment*)))

Filter for English language

Where available, tick the school age filter (6-12).

Where no ‘school age’ filter (ERIC), filter for grades 1-7, elementary, primary and exclude all others.

Where no appropriate filters available (Web of Science, Pro-Quest), add AND (primary OR elementary)

Appendix B Qualitative Quality Assurance

CASP

Criterion	Hamilton and Jones (2016)	Jones & Myhill (2010)
Section A: Are the results valid?		
1. Was there a clear statement of the aims of the research?	✓	✓
2. Is a qualitative methodology appropriate?	✓	✓

3. Is it worth continuing?	✓	✓
4. Was the recruitment strategy appropriate to the aims of the research?	✓	✓
5. Was the data collected in a way that addressed the research issue?	✓	✓
6. Has the relationship between researcher and participants been adequately considered?	✓	?
Section B: What are the results?		
7. Have ethical issues been taken into consideration?	✓	?
8. Was the data analysis sufficiently rigorous?	✓	X
9. Is there a clear statement of findings?	✓	✓
Section C: Will the results help locally?		
10. How valuable is the research?	?	✓

The University of Manchester

Appendix C Quantitative Quality Assurance

D.Ed.Ch.Psychol. 2017

Review framework for quantitative investigation research

Criterion	Campbell (2015)	Cline & Ertubey (1997)	Doherty & Hier (1988)	Hartas (2018)	Jones & Myhill (2010)	Plewis (1997)
Data gathering						
Clear research question or hypothesis *e.g. well-defined, measurable constituent elements*	✓	✓	✓	✓	X	✓
Appropriate participant sampling *e.g. fit to research question, representativeness.*	✓	✓	✓	✓	✓	✓

Appropriate measurement instrumentation. *e.g. sensitivity; specificity*	✓	✓	X	✓	?	✓
Comprehensive data gathering *e.g. multiple measures used; context of measurement recorded (e.g. when at school vs at home)*	✓	✓	✓	✓	✓	✓
Appropriate data gathering method used *e.g. soundness of administration*	✓	✓	✓	X	✓	?
Reduction of bias within participant recruitment/ instrumentation/ administration *e.g. harder-to-reach facilitation; accessibility of instrumentation*	X	X	X	✓	?	X

Response rate/ completion maximised *e.g. response rate specified; piloting; access options*	n/a	✓	✓	n/a	n/a	n/a
Population subgroup data collected *e.g. participant gender; age; location*	✓	✓	✓	✓	✓.	✓
Data analysis						
Missing data analysis *e.g. Level and treatment specified*	✓	n/a	n/a	n/a	n/a	n/a
Time trends identified *e.g. year on year changes*	n/a	✓	n/a	n/a	n/a	n/a
Geographic considerations *e.g. regional or subgroup analyses*	X	X	n/a	X	X	X

Appropriate statistical analyses (descriptive or inferential) *e.g. coherent approach specified; sample size justification.*	✓	✓	✓	✓	✓	✓
Multi-level or inter-group analyses present *e.g. comparison between participant groups by relevant location or characteristics*	✓	✓	✓	✓	✓	✓
Data interpretation						
Clear criteria for rating of findings *e.g. benchmarked/ justified evaluation of found quantitative facts*	✓	✓	✓	✓	X	✓
Limitations of the research considered in relation to initial aims *e.g. critique of method; generalizability estimate*	✓	✓	✓	✓	X	✓

Implications of findings linked to rationale of research question *e.g. implications for theory, practice or future research*	✓	✓	✓	✓	✓	✓

References

Choi, B.C.K. (1998). Perspectives on epidemiological surveillance in the 21st century. *Chronic Diseases in Canada, 19*(4), 145-151.

Cohen, L. , Manion, L., & Morrison, K. (2007) *Research Methods in Education (6th edition).* London: Routledge.

Geneady, A.M., Lemasters, G.K., Lockey, J., Succop, P., Deddens, J., Sobeih, T., & Dunning, K. (2007). An epidemiological appraisal instrument – a tool for evaluation of epidemiological studies. *Ergonomics, 50*(6), 920-960.

Wallace, M. & Wray, A. (2011). *Critical Reading and Writing for Postgraduates (2nd edition)*. London: Sage Publications.

Appendix D Data Extraction Table

What is the influence of pupil gender on teacher judgements and expectations of pupil outcomes?

Author	Title	Aim	Design (Qual / Quant / RCT etc. Method	Inc and exc criteria When and where conducted	Number of participants	Ages Gender	Measures	Type of analysis	Outcomes and key findings in relation to my review question
Campbell (2015)	Stereotyped at Seven? Biases in Teacher Judgement of Pupils' Ability and Attainment	To investigate "whether teacher-level stereotyping of pupils may relate to biased assessment according to pupil characteristics…" To explore "whether there are biases in teacher judgements of pupils which correspond to each of the key pupil characteristics underpinning	Quantitative Cognitive test of children (BAS word reading, shortened progress in mathematics test) Survey of teachers Lag time between tests	Inc. State school children in England. "MCS children who were still responding at wave four, whose teachers responded and where there is all necessary data."	Millennium cohort study participants Reading 4997 Maths 4985	7 and in year two Parent reported Reading boys 2494 Maths boys 2491 Reading girls 2503	Cognitive test of children (BAS word reading, shortened progress in mathematics test) Teacher survey (well above average, above average,	Regression modelling Linear probability model	These are a reflection of biased / inaccurate judgements. Boys are 4% less likely to be judged as above average in reading Boys are 5% more likely to be judged below average in reading Boys are 5.2% more likely to be judged above average in maths.

		recorded primary-age attainment gaps (family income level, gender, SEN, ethnicity, EAL)…"	(mean) was 3.8 months	Exc. Twins and triplets		Maths girls 2494	below average, well below average)		Boys are 2.1% more likely to be judged below average in maths
Cline & Ertubey (1997)	The impact of gender on primary teachers' evaluations of children's difficulties in school	"The aim of the study was to investigate gender influences on teachers' perceptions of children's difficulties across a different range of problematic aspects of development (learning as well as emotional difficulties)…"	Quantitative Vignettes then a questionnaire with a rating task (options for possible action that were hierarchically sequenced)	London, the home counties (e.g., Essex, Kent, Surrey, Hertfordshire) and North-West England	79 primary schools provided 523 questionnaires from teachers	Unknown (but length of experience given) Female n=442 (84.5%) Male n= 76 (14.5%) No information n=5 (0.9%)	6-point rating scale (Likert)	Cronbach-Alpha Reliability Coefficient ANOVA	"The overall results did not support the hypothesis that teachers' judgments on the need for external support would be influenced by the gender of the child concerned."
Doherty & Hier (1988)	Teacher Expectations and Specific Judgements: a small-scale	to test, and possibly extend, the	Quantitative Two independent	Junior school in Coventry.	5 teachers 113 children	Unknown but had "considerable teaching experience"	The Schonell Silent Reading Test (Form B)		Reading "a significant interaction effect involving Sex and

	study of the effects of certain non-cognitive variables on teachers' academic predictions	findings reported by Doherty & Conolly (1985) which were 1. teachers tended to over-estimate the scores in mathematics and English, and to under-estimate the scores in reading 2. academic competence was the single most powerful predictor of the primary teachers' estimations of their pupils' scores in mathematics, English and reading 3. sex and tidiness played an important part in influencing teachers' estimations of children's scores in	variables were sex and teacher's perception of child score.	Average class size 23		and were teaching year 3 and year 4 Adults all female Children: 64 boys, 57 girls	The Vernon and Miller's Graded Arithmetic-Mathematics Test The Piers-Harris Children's Self-Concept Scale The Barner-Lunn Academic Self-image Scale Rating scale for eight bi-polar constructs chosen by researchers and called a 'Personal Rating Scale' Regression analysis ANOVA	Personal Rating Scales." Pg 339 "the effect of teacher perception on predicted scores in reading is much more powerful for boys than for girls (although significant in both cases)." Pg 341 Maths "still a significant main effect for sex and for Personal Rating Scale, with academic competence controlled." Pg 342 "even with academic competence held constant, boys still receive a significantly higher predicted mean score." Pg 342 boys (who are less favourably perceived) are

		the three areas of attainment.							particularly "subject to faulty predictions" by teachers, especially in reading.
Hamilton & Jones (2016)	Illuminating the 'boy problem' from children's and teachers' perspectives: a pilot study	"What do boys think about their learning and school experiences?" "This study gives a voice to male pupils, aged six to seven years…" "…to illuminate the 'boy problem' from children's and teachers' perspectives." "To explore boys' perceptions of learning and schooling. To gain an insight into teacher perceptions of boys and their engagement in	Qualitative-interpretive-deductive paradigm, multiple methods of inquiry within case study design Semi structured interviews with teachers Discussion groups with boys (5 x 2 boys, 1 x 3 boys, 1 x 3 boys who had expressed dissatisfaction	Yr 2 State primary school in North Wales	13 children (6-7 yr old boys) 2 teachers Adults female (1 x experienced, 1 x recently qualified)		N/A	Narrative (based on Thomas, 2006)	"Extreme gender binaries held by teachers in relation: brain based and gendered learning patterns; boys' difficult behaviour and the use of 'syndromes' and 'labels'; the apparent difficulty some boys have with literacy; concerns about the impact that a 'feminised' education system can have on male learners " Pg 249 More tolerant of girls Belief that boys have distinct learning style to girls –

		schooling and learning."	with school in previous group) Unsystematic classroom observations						explaining why boys were in low ability literacy group and adapting feminist curriculum.
Hartas (2018)	Setting for English and maths: 11-year-olds' characteristics and teacher perceptions of school attitudes	to examine associations between set positions for English and Maths and 11-year-olds' psychosocial and cognitive characteristics and background factors and teachers' perceptions of their behaviour, attitude and post-16 educational goals. Are 11-year-olds more likely to be	Quantitative Home based surveys Home based interviews Teacher surveys	Millennium Cohort Study, fifth sweep 69% (2012 – 2013) Inc: Singleton children England and Wales	9610	11 50.9% boys 49.1% girls	Strengths and Difficulties Questionnaire Cambridge Gambling Task Home based surveys Home based interviews	Exploratory Factor Analysis with varimax rotation KMO statistic, Bartlet's test of sphericity, chi square, Hosmer Lemeshow	Boys 50% more likely than girls to be in the middle/lowest sets for English Boys 39% more likely than girls to be in the middle/lowest sets for Maths "Teachers' expectations regarding post-16 education and ratings of child behaviour and attitudes, family income, **gender** and children's decision-making emerged as strong

		placed in a middle or lower set depending on whether they are male or female?					Teacher surveys	Binary logistic regression analysis Linear regression Negelkerke test for effect size Wald statistic	predictors, [of set position]…" "This suggests that perceived attitudinal gender differences may be used by teachers to explain attainment differences."
Jones & Myhill (2010)	'Troublesome boys' and 'compliant girls': gender identity and perceptions of achievement and underachievement	To explore the relationship between teacher's perceptions of gender and achievement.	Qualitative with some quantitative observation data Semi-structured interviews of 36 class teachers, all English Dept teachers in high school, 4 teachers from other	Project JUDE-South-West England 10 self-selecting schools 6 x first schools, 3 x middle schools, 1 x high school	40 teachers 144 children Classes: 6 x yr 1 (age 5-6) 6 x yr 4 (age 8-9) 6 x yr 5 (age 9-10) 6 x yr 8 (age 12-13) 6 x yr 9 (age 13-14) 6 x yr 10 (age 14-15)		Off task scores and ranks designed for this study	Tally of comments ANOVA	"Underachieving boys … were viewed as typical boys." "Comments about high-achieving girls tended to say that they were typical." "The representation of boys as active rather than passive, and as challenging rather than accepting, is seen to contribute to their underachievement." "Stereotypical gender identities persist, in spite of

			curriculum subjects Interview and observation of 4 specific children per class selected by class teacher		Gender unknown but subgroup information collected about gender and achievement levels			individuals who clearly do not conform to gender expectations. Teachers have formed a strong set of concepts and opinions in relation to the underachieving boy.”
Plewis (1997)	Inferences about teacher expectations from national assessment at key stage one	to use KS1 assessment data “to make inferences … about the way teachers systematically perceive, and make judgments about, pupils from different socio-demographic groups.”	Quantitative Analysis of SAT and Teacher Assessment data at KS1	First data set: 16 self-selected local education authorities Second data set: 70 local education authorities Evaluation of National Curriculum Assessment team (ENCA)	First data set: 80 schools 2400 pupils Second data set: 250 schools 5000 pupils 6-7 yrs old Gender unknown	SAT TA	Kappa coefficient	“statistically significant effect for gender for English. Girls are 1.15 times as likely as boys to get a TA rating of two rather than three relative to their SAT level.” No significant effects seen here for maths or science. a consistent pattern across English, maths and science which translates into the odds of a girl getting a higher TA for a fixed SAT level being

				National Foundation for Educational Research (NFER)				about 20 per cent greater than for a boy.

Appendix E OPTIC

Observing Pupils and Teachers In Classrooms (OPTIC) Schedule

School ______________________________ Class __________

Teacher ______________________________ Date __________ Time __________

No. of children present __________ Observer 1 __________ Observer 2 __________

Nature of lesson ______________________________

Section A. Teacher behaviours

Positive responses		Negative responses	
Academic	Social	Academic	Social
1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10	1 2 3 4 5 6 7 8 9 10

Totals (REX)

( )	( )	( )	( )

Section B. Pupil behaviours

Total = per cent on-task

© Frank Merrett & Kevin Wheldall, 1986
All rights reserved

Appendix F Categorisation of comments

Comments refer to past events only

Positive comments about academic work or the task (implying the teacher wants MORE of this)	Redirections or disapproving comments about academic work or the task (implying the teacher wants LESS of this)	Positive comments about social skills or behaviour (implying the teacher wants MORE of this)	Redirections or disapproving comments about social skills or behaviour (implying the teacher wants LESS of this)
Correct That's great! I like that! Well done! OK (positive tone) Not bad! Super duper Wowzers That's it! You're on fire Repetition of an answer to validate and confirm	That's wrong, You're a dead loss, Don't do that! You haven't edited Go and read it again. You have a habit of rushing things Have another think It's your time you are wasting (when waiting for work completion). What have you forgotten? Individual feedback on work which includes recognising , criticising, or redirecting mistakes	I'm pleased to see you working quietly Well done for … Child x is sitting beautifully Lovely lining up Great listening Not bad! Super duper Wowzers That's really kind Thank you	Look at John, everyone. This is what I call a really silly boy Shhh Child's name (with negative tone intended to stop behaviour) Don't just copy If you weren't talking, we could be out to play by now. It's your time you are wasting (when waiting for quiet). Why are you (still) talking / standing up? Looking this way *name*

Individual feedback on work which includes acknowledgement of correct answers or good completion of the task.		Individual feedback which includes acknowledgement of good learning behaviours e.g., effort, perseverance Positive reference to seeing school values or golden rules	Excuse me (tone implies negativity) Individual feedback which redirects learning behaviours (e.g., not listening, lack of effort) or social behaviours (e.g., talking, wandering around, arguing, not taking turns or sharing)

NB: Comments in blue from Merrett & Wheldall (1986)

Not included in data collection	
Guidance or instructions for the future	Next time try to include more adjectives
Instructions on how to complete a learning task	Remember to write the date. I'm expecting 4 paragraphs.
Generic questions	Are you OK?
Double comments counted as one (Well done – you were right)	

Appendix G Topic Guide

UNIVERSITY OF Southampton

Project title: Teacher goals and the distribution, content and self-perception of teacher language across contexts

ERGO number: 70556

Interview Questions:

- The year group of your class
- The number of children in the class
- The number of children in your small group
- How many years you have been in teaching

1. Can you place these 20 unifix in the 4 boxes represented in the table below, thinking about the number of comments made during the whole class session.
    i. 1 unifix = 5%
    ii. 2 unifix = 10%
    iii. 4 unifix = 20%
    iv. 6 unifix = 30%
    v. 8 unifix = 40%
    vi. 10 unifix = 50%
    vii. 12 unifix = 60%
    viii. 14 unifix = 70%
    ix. 16 unifix = 80%
    x. 18 unifix = 90%
    xi. 20 unifix = 100%

Positive comments about academic work or the task	Redirections or disapproving comments about academic work or the task	Positive comments about social skills or behaviour	Redirections or disapproving comments about social skills or behaviour

2. Please explain why you have placed them as you have?
3. Repeat unifix task, and explanation question for the small group session

SHARE DATA FROM OBSERVATION INCLUDING WHETHER THEY ARE ABOVE OR BELOW THE EXISITING RESEARCH MEAN FOR POSITIVE / APPROVING COMMENTS AND FOR REDIRECTIONS / DISAPPROVING COMMENTS

4. Can you tell me your thoughts on the data I have just shown you? How does this match with your predictions and hopes? What are your thoughts about the previous research figures?
5. Please can you tell me what you were hoping for at the beginning of the lesson.
    a. Firstly, what did you hope the children would achieve academically? Did you feel the lesson would challenge them, help them consolidate learning or generally be easy to grasp?
    b. Now, what were your expectations for the children with regard to their social behaviour? For example, were you hoping for independence, co-operation, resilience particularly?
    c. Were these expectations specific in any way to this particular lesson?
6. Reflect on a short interaction you remember having with a pupil during whole class teaching where you feel the pupil was meeting your hopes and expectations.
    a. Describe what was happening in the room.
    b. Tell me what you were thinking and what action you took (including what you may have said).
    c. Explain what outcome you were hoping for and why.
7. Ask for a second example but which took place in the small group

8. Reflect on a short interaction you remember having with a pupil during whole class teaching where you feel the pupil was NOT meeting your hopes and expectations or perhaps you found difficult.
    a. Describe what was happening in the room.
    b. Tell me what you were thinking and what action you took (including what you may have said).
    c. Explain what outcome you were hoping for and why.
9. Ask for a second example as above
10. The purpose of this research is to consider teachers' goals and the way teachers use and perceive they use language in the classroom (both small group and whole class). Is there anything else you'd like to add about this topic that we haven't discussed?

Appendix H Ethical approval documents

ERGO II Ethics application form – Psychology Committee

1. Applicant Details

1.1 Applicant name	**Mrs Rebecca Thompson**
1.2 Supervisor	**Dr Tim Cooke, Dr Fiona Okai**
1.3 Other researchers / collaborators (if applicable): *Name, address, email*	

2. Study Details

2.1 Title of study	**Teacher goals and the distribution, content and self-perception of teacher language across contexts**
2.2 Type of project (e.g. undergraduate, Masters, Doctorate, staff)	**Doctoral thesis**

2.3 Briefly describe the rationale for carrying out this project and its specific aims and objectives.

The most common use of teacher language is to make positive comments about academic tasks, followed by criticism of behaviour. Positive comments regarding behaviour are less frequent and lastly is the use of criticism of academic work. However, there is currently a paucity of research into how language might change within the same classroom but between the teacher in a whole class context and when they are working with a small group.

Apter et al. (2010) found that frequent use of positive feedback about academic work was linked with high percentage rates of student on-task behaviour during lessons. In subsequent research (Apter et al., 2020) found, when comparing rates of verbal feedback in English and maths lessons, that the English department were using high numbers of positive comments about task performance and achieving the highest percentage of 'on-task' behaviours from pupils. Similarly, research has shown that when positive to reprimand ratios increase, the 'on-task' behaviour of the whole class improves (Caldarella et al., 2021). This evidence suggests a clear link between language distribution patterns and pupil behaviour. What these studies do not show however, is why teachers use language in the way they do.

When considering this question, it is important to understand what might be influencing teachers' direction and use of language. Teacher attitudes were considered in the Pygmalion in the classroom study (Rosenthal & Jacobson, 1968) and suggested that teacher expectations could have an important impact on the outcomes of pupils; the beliefs of teachers (even if false) have the potential to create a self-fulfilling prophecy. In the subsequent plethora of expectations research, including very recent studies demonstrating that teacher expectations positively predict pupil outcomes (Tandler & Dalbert, 2020; Wang et al., 2021) there was evidence that some teachers interact differently with pupils depending on perceived capability, but others did not (Good et al., 2018).

Priorities of teachers in decision making must also be considered. Pirskanen et al. (2019) found that teachers of young children prioritised happiness and excitement in school alongside the teaching of emotion management. Academic skills were felt to be of secondary importance, developing later in the year.

It is therefore important not just to gather statistical information about the type and quantity of language teachers use, but also what role their attitude and intentions may have in these decisions.

It is hoped that this study will help improve the current understanding of language used in classrooms and why this is the case. This will help researchers to:

- provide important information for coaching and mentoring teachers.
- consider how certain contexts shape the language teachers use thus informing pedagogical practice
- provide important information for supporting small group intervention packages

2.4 Provide a brief outline of the basic study design. Outline what approach is being used and why.

This will be a mixed methods approach using observation of classes and groups, followed by semi-structured, individual interviews of teachers. The observations will be documented using part one of the Observing Pupils and Teachers In Classrooms (OPTIC) schedule and the interviews will be video recorded and transcribed. Transcripts will be analysed using narrative analysis.

2.5 What are the key research question(s)? Specify hypotheses if applicable.

What is the distribution of academic praise and academic criticism by primary school teachers across different contexts (group and whole class)?

What is the distribution of social behaviour praise and social behaviour criticism by primary school teachers across different contexts (group and whole class)?

How well calibrated are primary school teacher predictions and perceptions of their use of language to observational data?

How do primary school teachers articulate what they are trying to achieve when using different types of language?

Are there qualitative topics which occur frequently in primary teacher narratives?

3. Sample and setting

3.1 Who are the proposed participants and where are they from (e.g. fellow students, club members)? List inclusion / exclusion criteria if applicable.
Participants will be primary school teachers who will teach their own class for the observation. The same teacher will be observed in the small group context but still within the classroom setting. Teachers who cover planning, preparation and assessment time (PPA), cover for absences or are specialist (e.g., music) teachers will not be used. This will ensure that observed participants have enduring knowledge and understanding of the individual children in their class and have a responsibility for each child. Choice of school will be limited to those in which teacher led groups are common practice. I will be guided by Headteachers and SENCos as to which teachers fit the inclusion criteria (e.g., avoiding those on performance management who may be receiving increased observations already) but I will offer guidance that ideally a diverse (experience, gender, age, ethnicity) range of teachers would be preferable.

3.2. How will the participants be identified and approached? Provide an indication of your sample size. If participants are under the responsibility of others (e.g., parents/carers, teachers) state if you have permission or how you will obtain permission from the third party).
I have contacted Headteachers at schools I know through previous contact or through my work as a TEP. I currently have initial expressions of interest from 5 headteachers to conduct research in their schools. I will use email to stay in contact with the school and make arrangements. In the XXX email I state that I will submit to the ethics committee in the Spring Term and wait for approval. In the XXX and XXX emails I state that I will submit to the ethics committee and wait for approval. In the XXX email I state that I am currently putting together an ethics application and official documentation, and will be in touch as things move forward. In the XXX email I state that it is an initial outline.

3.3 Describe the relationship between researcher and sample. Describe any relationship e.g., teacher, friend, boss, clinician, etc.
Gatekeepers Two of the headteachers were XXX. One of the headteachers is XXX One of the headteachers is XX One of the headteachers is XXX Participants There is a possibility that some participants XXX

3.4 How will you obtain the consent of participants? (*please upload a copy of the consent form if obtaining written consent***) NB A separate consent form is not needed for online surveys where consent can be indicated by ticking/checking a consent box (normally at the end of the PIS). Other online study designs may still require a consent form or alternative procedure (for example, recorded verbal consent for online interviews).**
I will offer a video explanation of my thesis or an in person question-and-answer session (school's choice). Following this I will ask for signed consent forms.

3.5 Is there any reason to believe participants may not be able to give full informed consent? If yes, what steps do you propose to take to safeguard their interests?
No

4. Research procedures, interventions and measurements

4.1 Give a brief account of the procedure as experienced by the participant. Make it clear who does what, how many times and in what order. Make clear the role of all assistants and collaborators. Make clear the total demands made on participants, including time and travel. *Upload copies of questionnaires and interview schedules to ERGO.*
The teacher will be asked to teach their class for an hour as they normally would (ensuring both whole class teaching and at least 20 minutes of small group teaching). I will observe the lesson and document (hard copy) the language used. For some data collection, a Voluntary Research Assistant (VRA) named XXX will be asked to join me for the purposes of checking inter-rater reliability. During the day on which I conduct the observation, I will conduct a semi-structured interview for between 60 and 90 minutes. The interviews will be video recorded on a University of Southampton laptop for later transcript and analysis.

4.2 Will the procedure involve deception of any sort? If yes, what is your justification?
No

4.3. Detail any possible (psychological or physical) discomfort, inconvenience, or distress that participants may experience, including after the study, and what precautions will be taken to minimise these risks.
Teachers may feel compelled to participate if their headteacher has volunteered the school or the school needs the money for the PTA. I will make clear on the Participant Information Sheet that payments are not based on the numbers of participating teachers and I will not refuse money if teachers withdraw from the study. Teachers will be reassured that their data is not for any performance management or appraisal processes. Individual data will not be shared with their headteacher. A second concern may be if there are wide disparities between teacher perception and observational data which is perceived as wrong or negative by the teacher. (For example

if the teacher believes they are someone who uses lots of positive praise and generates a positive atmosphere in the classroom, they may feel discomfort or cognitive dissonance if my initial statistics show differently.) This will be addressed through thorough, verbal debriefing sessions and a reiteration that observations are a 'snapshot' rather than an in-depth analysis of one teacher.

4.4 Detail any possible (psychological or physical) discomfort, inconvenience, or distress that YOU as a researcher may experience, including after the study, and what precautions will be taken to minimise these risks. If the study involves lone working please state the risks and the procedures put in place to minimise these risks (<u>please refer to the lone working policy</u>).
None

4.5 Explain how you will care for any participants in 'special groups' e.g., those in a dependent relationship, are vulnerable or are lacking mental capacity), if applicable:
n/a

4.6 Please give details of any payments or incentives being used to recruit participants, if applicable:
£100 to the parent teacher association (PTA) of each school

5. Access and storage of data

5.1 How will participant confidentiality be maintained? Confidentiality is defined as non-disclosure of research information except to another authorised person. Confidential information can be shared with those already party to it and may also be disclosed where the person providing the information provides explicit consent. Consider whether it is truly possible to maintain a participant's involvement in the study confidential, e.g. can people observe the participant taking part in the study? How will data be anonymised to ensure participants' confidentiality?
It may be possible to see that a teacher is being observed or interviewed through internal windows but no individual data will be shared. In the unlikely circumstances that audio and video recordings will be made (e.g. if COVID prevents access to classrooms) the school will be asked to use their own recording equipment and transfer files to the University of Southampton system using TeamViewer which encrypts files at both ends and is already used by UoS. I will allocate pseudonyms (names rather than codes) to participants and all data collected (OPTIC and interview) will be stored using the pseudonym. An anonymization key will be kept securely on the University system.

5.2 How will personal data and study results be stored securely during and after the study. Who will have access to these data?
All data (e.g., OPTIC results / videos and transcriptions of the interviews) will be held securely and in line with General Data Protection Regulations (2018) using the University of Southampton's IT system. File size is not expected to be excessive (less than 1GB). Hard copies of paperwork will be scanned and destroyed. Demographic information about the age and gender of teachers will not be collected, but years of experience in the classroom will be. After the study, data will be committed to the archive service who will maintain it for a minimum of 10 years as per the University RDM Policy.

5.3 How will it be made clear to participants that they may withdraw consent to participate? Please note that anonymous data (e.g. anonymous questionnaires) cannot be withdrawn after they have been submitted. If there is a point up to which data can be withdrawn/destroyed e.g., up to interview data being transcribed please state this here.
I will offer a video explanation of my thesis or a question-and-answer session (school's choice). Following this I will ask for signed consent forms which will clearly state the right to withdraw up to interview data being transcribed.

6. Additional Ethical considerations

6.1 Are there any additional ethical considerations or other information you feel may be relevant to this study?
Some teaching unions recommend that teachers are only observed by people with Qualified Teacher Status. I do hold this qualification so will include this in my participant information sheet and will adhere to the NASUWT protocol for classroom observation. If any child with whom my participants are interacting is familiar to me, I will not discuss this child with anyone outside of the research team. I may discuss the relevant child with the participant during the interview process but not beyond this. As a TEP I hold DBS obtained through the University. When I use a Voluntary Research Assistant, I will ensure they too hold a DBS. VRAs who do not hold DBS will be allowed to transcribe but not participate in data collection in schools.

CONSENT FORM

Study title: **Teacher goals and the distribution, content and self-perception of teacher language across contexts**

Researcher name: **Rebecca Thompson**

ERGO number: 70556

Participant Identification Number:

Please initial the box(es) if you agree with the statement(s):

I have read and understood the information sheet *Version 1, dated 5.3.22* and have had the opportunity to ask questions about the study.	
I agree to take part in this research project and agree for my data to be used for the purpose of this study.	
I agree to take part in the ***observation*** for the purposes set out in the participation information sheet and understand that this will be recorded using written notes.	
I agree to take part in the ***interview*** for the purposes set out in the participation information sheet and understand that this will be recorded using video.	
I understand my participation is voluntary and I may withdraw up until my observation data or interview has been analysed (whichever is sooner) for any reason without my participation rights being affected. I understand that my school will still receive the promised funding if I withdraw.	
I understand that I may be quoted directly in reports of the research but that I will not be directly identified (e.g., that my name will not be used and that no specifically identifiable references will be made to my school).	

I give permission for the data and interview comments that I provide to be held by the University of Southampton archive service as described in the participant information sheet so it can be used for future research and learning in the area of pedagogy.	

Name of participant (print name) ……

Signature of participant……

Date……

Name of researcher (print name) …Rebecca Thompson

Signature of researcher ……

Date……

Participant Information Sheet

Study Title: **Teacher goals and the distribution, content and self-perception of teacher language across contexts**

Researcher: Rebecca Thompson

ERGO number: 70556

You are being invited to take part in the above research study. To help you decide whether you would like to take part or not, it is important that you understand why the research is being done and what it will involve. Please read the information below carefully and ask questions if anything is not clear or you would like more information before you decide to take part in this research. You may like to discuss it with others but it is up to you to decide whether or not to take part. If you are happy to participate you will be asked to sign a consent form.

What is the research about?

I am working towards a Doctorate in Educational Psychology and am conducting research for a thesis. I have worked in education for many years as a class teacher (I hold qualified teacher status) and then a SENCo. This research continues my interest in classroom pedagogy and the difference a teacher can make. This particular research explores how teacher attitudes and intentions impact teacher use of language within the classroom and how this changes in different contexts such as a small group. I am hoping the research will be helpful in terms of coaching and mentoring teachers and informing pedagogical practice. I am interested in gathering information about:

- *What is the distribution of academic praise and academic criticism / social behaviour praise and social behaviour criticism across different contexts (group and whole class)?*
- *How well calibrated are teacher predictions and perceptions of their use of language to observational data?*
- *How do teachers articulate what they are trying to achieve when using different types of language?*

Why have I been asked to participate?

I am observing 12 teachers across 4 schools plus conducting a pilot study in a 5th school. I have asked head teachers to invite teachers who teach across a variety of age ranges in their schools to enable me to gather a broad range of data.

What will happen to me if I take part?

I will come into your class on a day agreed by us and your school. I will observe you for one hour, making notes and observations about the different language you use. I may bring a research assistant with me who will also observe and take notes.

I would like you to spend at least 20 consecutive minutes during the hour working with a small group within the classroom itself. Please do not select specific children for my benefit. Plan as you usually would.

Following the observation, I will spend 60 – 90 minutes interviewing you about the lesson. The interview will focus on your language. (I will not be asking about curriculum or assessment). I will video record the interview in order to transcribe and analyse it afterwards.

The thesis is due to be completed by June 2023. I will send out a summary of the research findings during the autumn of 2023.

Are there any benefits in my taking part?

Participating in this study will help improve our current understanding of language used in classrooms. This will help researchers to:

- provide important information for coaching and mentoring teachers.
- consider how certain contexts shape the language teachers use thus informing pedagogical practice
- provide important information for supporting small group intervention packages

Every participating school will receive £100 for their PTA fund. This benefit will still be given to the school even if individual teachers decide to withdraw from the research.

Thank you for participating

Are there any risks involved?

No

What data will be collected?

I will collect:

- The year group of your class
- The number of children in the class
- The number of children in your small group
- The amount of time (in minutes) I observe you
- How many years you have been in teaching
- A transcript from your interview
- A video of your interview

Hard copies of information will be locked into a workbag for transport. All data will then be transferred to the University of Southampton laptop and IT system. Hard copies will be destroyed once inputted. Electronic data (e.g., scanned paperwork, video recordings, transcriptions) will be held securely and in line with General Data Protection Regulations (2018) using the University of Southampton's IT system.

Contact information is held in the University of Southampton password protected email system. Microsoft authenticator is also required for access.

Will my participation be confidential?

We will handle all data confidentially. At the end of the research, findings will be presented as a whole, and individual data will not be identifiable. Any names of children, schools, classes or other information that may make you personally identifiable (e.g., "I was off sick with COVID for 6 weeks last term,") will be removed from the transcription. However, it is likely that others in your school will know that you are taking part.

Only members of the research team and responsible members of the University of Southampton (including voluntary research assistants) may be given access to data about you for monitoring purposes and/or to carry out an audit of the study to ensure that the research is complying with applicable regulations. Individuals from regulatory authorities (people who check that we are carrying out the study correctly) may require access to your data. All of these people have a duty to keep your information, as a research participant, strictly confidential.

Once your interview is transcribed, the original recording will be destroyed. There will be an anonymization key to link the classroom data to the transcription.

Do I have to take part?

No, it is entirely up to you to decide whether or not to take part. If you decide you want to take part, you will need to sign a consent form to show you have agreed to take part.

What happens if I change my mind?

You have the right to change your mind and withdraw at any time without giving a reason and without your participant rights being affected. If you wish to withdraw, please email r.s.thompson@soton.ac.uk. You may withdraw up until your interview is transcribed.

What will happen to the results of the research?

Your personal details will remain strictly confidential. Research findings made available in any reports or publications will not include information that can directly identify you without your specific consent.

You will receive a summary copy of the results and findings and a copy will be sent to your school. The research will be written up in full into a thesis submission. You can request a copy of the full thesis after October 2023. It may also be sent for publication in a journal.

After the report has been written up, the pseudonymised data will be stored on the University of Southampton's Data Repository archive system. It is anticipated that the anonymised research data may be made accessible to the wider academic community in order to support further research in the area. In this way future research could also aim to support better pedagogical practice.The access to this data will not include commercial use.

If you wish to file a complaint, please follow the instructions below.

Where can I get more information?

If you have any further questions, please email r.s.thompson@soton.ac.uk or edpsych-fels@soton.ac.uk

What happens if there is a problem?

If you have a concern about any aspect of this study, you should speak to the researchers who will do their best to answer your questions.

If you remain unhappy or have a complaint about any aspect of this study, please contact the University of Southampton Research Integrity and Governance Manager (023 8059 5058, rgoinfo@soton.ac.uk).

Contact information

Rebecca Thompson r.s.thompson@soton.ac.uk

Dr Tim Cooke t.cooke@soton.ac.uk

Dr Fiona Okai F.Okai@soton.ac.uk

Faculty edpsych-fels@soton.ac.uk

Data Protection Privacy Notice

The University of Southampton conducts research to the highest standards of research integrity. As a publicly-funded organisation, the University has to ensure that it is in the public interest when we use personally-identifiable information about people who have agreed to take part in research. This means that when you agree to take part in a research study, we will use information about you in the ways needed, and for the purposes specified, to conduct and complete the research project. Under data protection law, 'Personal data' means any information that relates to and is capable of identifying a living individual. The University's data protection policy governing the use of personal data by the University can be found on its website (https://www.southampton.ac.uk/legalservices/what-we-do/data-protection-and-foi.page).

This Participant Information Sheet tells you what data will be collected for this project and whether this includes any personal data. Please ask the research team if you have any questions or are unclear what data is being collected about you.

Our privacy notice for research participants provides more information on how the University of Southampton collects and uses your personal data when you take part in one of our research projects and can be found at http://www.southampton.ac.uk/assets/sharepoint/intranet/ls/Public/Research%20and%20Integrity%20Privacy%20Notice/Privacy%20Notice%20for%20Research%20Participants.pdf

Any personal data we collect in this study will be used only for the purposes of carrying out our research and will be handled according to the University's policies in line with data protection law. If any personal data is used from which you can be identified directly, it will not be disclosed to anyone else without your consent unless the University of Southampton is required by law to disclose it.

Data protection law requires us to have a valid legal reason ('lawful basis') to process and use your Personal data. The lawful basis for processing personal information in this research study is for the performance of a task carried out in the public interest. Personal data collected for research will not be used for any other purpose.

For the purposes of data protection law, the University of Southampton is the 'Data Controller' for this study, which means that we are responsible for looking after your information and using it properly. The University of Southampton will keep identifiable information about you for 10 years after the study has finished after which time any link between you and your information will be removed.

To safeguard your rights, we will use the minimum personal data necessary to achieve our research study objectives. Your data protection rights – such as to access, change, or transfer such information - may be limited, however, in order for the research output to be reliable and accurate. The University will not do anything with your personal data that you would not reasonably expect.

If you have any questions about how your personal data is used, or wish to exercise any of your rights, please consult the University's data protection webpage (https://www.southampton.ac.uk/legalservices/what-we-do/data-protection-and-foi.page) where you can make a request using our online form. If you need further assistance, please contact the University's Data Protection Officer (data.protection@soton.ac.uk).

Data will be pseudonymised. Data that has been pseudonymised through key-coding and removal of personal identifiers still falls within the scope of the GDPR. This is because the data that allows identification of that person still exists, just not all in one place. Pseudonymised data can help reduce privacy risks by making it more difficult to identify individuals, but it is still personal data. If you are using pseudonymisation, i.e. linking data using a code, this should be explained with details on who can access the codes so as to enable an individual to be identified.

Thank you for taking the time to read this information sheet and consider being part of the research.

Appendix I Examples of coding from transcripts

P2: Well, I think um in - kind of - year one, because they young children, I think you have to do a lot more kind of positive um, reinforcement - sorry not reinforcement – ah social and academic um, with the children because they're so young, and that's the kind of jolly them along is kind of the way you - kind of -teach. Whether you'd see a contrast if you did it um, further up the school whether they would use less of that. Um, and particularly I think I do teach a lot like that. So, I do teach with a lot of positive comments, and I think probably the reason I scored myself lower in the two reinforcement ones, as I think you're less aware of how much you probably do the 'oh make sure you sitting properly' the/or 'make sure you're doing your um letters the right way, doing your finger spaces' cause I think that's all kind of - it's - it comes as more of an automatic thing. Whereas because you're trying to encourage the children, you're positive is more memorable if you know what I mean. So, I think I probably remember that more, so I thought that's why I would score higher in those areas.

Age or year group of the child important

Teacher feels they use a lot of positive language

Redirective language is more automatic

Using positive language is more memorable

P5: Well, I mean I would prefer to be more positive than I am disapproving, so it's a bit disappointing. The disapproving was, so high there.

Disappointment when redirecting scores are high / positive teacher comments are good or valuable

I: You were still... well I think that was just um, I did notice the lots of non-verbal. There was a particular boy whose name began with a [redacted], that you had to use his name quite a bit...

P5: Hmm, yes that's true ... He wouldn't pick up on non-verbal cues, right.

Individual children / Specific needs

I: Yeah, that's interesting, that's really interesting.

P5: Like, if I did that to him like, or I do some other general, try make eye contact, he wouldn't recognise that that was him that he needed to change what he was doing.

Individual children / Specific needs

I: So, you're kind of adapting your way of re-directing depending on the child. So, you know ...

P5: Yeah.

I: So that's how you know ...

P5: Yeh, I wouldn't do the same every child, no, no, he needs, he has to have it verbally.

Individual children / Specific needs

I: And you know him, and you've obviously been teaching him all year.

P5: Whereas other children, throughout the year have had less and less verbal redirections and reminders because they learned the non-verbal cues.

Verbal cues and decrease and non-verbal cues can increase / Variation across time

Appendix J ANOVA Statistics

	Group			Statistic	Std. Error
Comments per minute	Whole class	Mean		.5158	.05801
		95% Confidence Interval for Mean	Lower Bound	.3991	
			Upper Bound	.6325	
		5% Trimmed Mean		.4878	
		Median		.4650	
		Variance		.162	
		Std. Deviation		.40191	
		Minimum		.00	
		Maximum		1.63	
		Range		1.63	
		Interquartile Range		.55	
		Skewness		.900	.343
		Kurtosis		.500	.674
	Small group	Mean		.4879	.08425
		95% Confidence Interval for Mean	Lower Bound	.3184	
			Upper Bound	.6574	
		5% Trimmed Mean		.4306	
		Median		.2350	
		Variance		.341	
		Std. Deviation		.58369	
		Minimum		.00	
		Maximum		2.33	
		Range		2.33	
		Interquartile Range		.72	
		Skewness		1.514	.343
		Kurtosis		1.361	.674

	Comment type			Statistic	Std. Error
Comments per minute	Positive academic	Mean		1.1708	.09379
		95% Confidence Interval for Mean	Lower Bound	.9768	
			Upper Bound	1.3649	
		5% Trimmed Mean		1.1522	
		Median		1.1000	
		Variance		.211	
		Std. Deviation		.45949	
		Minimum		.39	
		Maximum		2.33	
		Range		1.94	
		Interquartile Range		.72	
		Skewness		.592	.472
		Kurtosis		.129	.918
	Positive social	Mean		.2138	.04026
		95% Confidence Interval for Mean	Lower Bound	.1305	
			Upper Bound	.2970	
		5% Trimmed Mean		.1979	
		Median		.1700	
		Variance		.039	
		Std. Deviation		.19722	
		Minimum		.00	
		Maximum		.72	
		Range		.72	
		Interquartile Range		.24	
		Skewness		1.224	.472
		Kurtosis		1.339	.918
	Redirect academic	Mean		.2021	.02920
		95% Confidence Interval for Mean	Lower Bound	.1417	
			Upper Bound	.2625	
		5% Trimmed Mean		.1928	
		Median		.2050	
		Variance		.020	
		Std. Deviation		.14307	
		Minimum		.00	

	Maximum		.60	
	Range		.60	
	Interquartile Range		.17	
	Skewness		.853	.472
	Kurtosis		1.268	.918
Redirect social	Mean		.4208	.06531
	95% Confidence Interval for Mean	Lower Bound	.2857	
		Upper Bound	.5559	
	5% Trimmed Mean		.4053	
	Median		.3550	
	Variance		.102	
	Std. Deviation		.31997	
	Minimum		.00	
	Maximum		1.13	
	Range		1.13	
	Interquartile Range		.52	
	Skewness		.648	.472
	Kurtosis		-.530	.918

	Group and type of comment			Statistic	Std. Error
Comments per minute	Whole class positive academic	Mean		.9417	.11068
		95% Confidence Interval for Mean	Lower Bound	.6981	
			Upper Bound	1.1853	
		5% Trimmed Mean		.9341	
		Median		.8250	
		Variance		.147	
		Std. Deviation		.38340	
		Minimum		.39	
		Maximum		1.63	
		Range		1.24	
		Interquartile Range		.67	
		Skewness		.719	.637
		Kurtosis		-.528	1.232
	Whole class positive social	Mean		.3267	.06081
		95% Confidence Interval for Mean	Lower Bound	.1928	
			Upper Bound	.4605	
		5% Trimmed Mean		.3230	
		Median		.2800	
		Variance		.044	
		Std. Deviation		.21064	
		Minimum		.00	
		Maximum		.72	
		Range		.72	
		Interquartile Range		.24	
		Skewness		.696	.637
		Kurtosis		.216	1.232
	Whole class redirect academic	Mean		.1517	.04719
		95% Confidence Interval for Mean	Lower Bound	.0478	
			Upper Bound	.2555	
		5% Trimmed Mean		.1352	
		Median		.1150	
		Variance		.027	
		Std. Deviation		.16348	
		Minimum		.00	
		Maximum		.60	
		Range		.60	
		Interquartile Range		.19	
		Skewness		2.013	.637
		Kurtosis		5.260	1.232
		Mean		.6433	.07790
			Lower Bound	.4719	

Whole class redirect social	95% Confidence Interval for Mean	Upper Bound	.8148	
	5% Trimmed Mean		.6387	
	Median		.5900	
	Variance		.073	
	Std. Deviation		.26986	
	Minimum		.24	
	Maximum		1.13	
	Range		.89	
	Interquartile Range		.35	
	Skewness		.311	.637
	Kurtosis		-.450	1.232
Small group positive academic	Mean		1.4000	.12243
	95% Confidence Interval for Mean	Lower Bound	1.1305	
		Upper Bound	1.6695	
	5% Trimmed Mean		1.3772	
	Median		1.3050	
	Variance		.180	
	Std. Deviation		.42411	
	Minimum		.88	
	Maximum		2.33	
	Range		1.45	
	Interquartile Range		.65	
	Skewness		.845	.637
	Kurtosis		.509	1.232
Small group positive social	Mean		.1008	.02759
	95% Confidence Interval for Mean	Lower Bound	.0401	
		Upper Bound	.1616	
	5% Trimmed Mean		.0959	
	Median		.0800	
	Variance		.009	
	Std. Deviation		.09558	
	Minimum		.00	
	Maximum		.29	
	Range		.29	
	Interquartile Range		.15	
	Skewness		.882	.637
	Kurtosis		-.019	1.232
Small Group redirect academic	Mean		.2525	.02962
	95% Confidence Interval for Mean	Lower Bound	.1873	
		Upper Bound	.3177	
	5% Trimmed Mean		.2528	
	Median		.2500	

		Variance		.011	
		Std. Deviation		.10261	
		Minimum		.07	
		Maximum		.43	
		Range		.36	
		Interquartile Range		.12	
		Skewness		-.100	.637
		Kurtosis		-.010	1.232
	Small group redirect academic	Mean		.1983	.05263
		95% Confidence Interval for Mean	Lower Bound	.0825	
			Upper Bound	.3142	
		5% Trimmed Mean		.1831	
		Median		.1300	
		Variance		.033	
		Std. Deviation		.18230	
		Minimum		.00	
		Maximum		.67	
		Range		.67	
		Interquartile Range		.20	
		Skewness		1.809	.637
		Kurtosis		3.515	1.232

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group and type of comment	Statistic	df	Sig.	Statistic	df	Sig.
Comments per minute	Whole class positive academic	.251	12	.035	.896	12	.142
	Whole class positive social	.160	12	.200*	.934	12	.421
	Whole class redirect academic	.243	12	.048	.780	12	.006
	Whole class redirect social	.147	12	.200*	.960	12	.788
	Small group positive academic	.173	12	.200*	.935	12	.434
	Small group positive social	.170	12	.200*	.900	12	.157
	Small group redirect academic	.138	12	.200*	.978	12	.974
	Small group redirect social	.313	12	.002	.792	12	.008

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

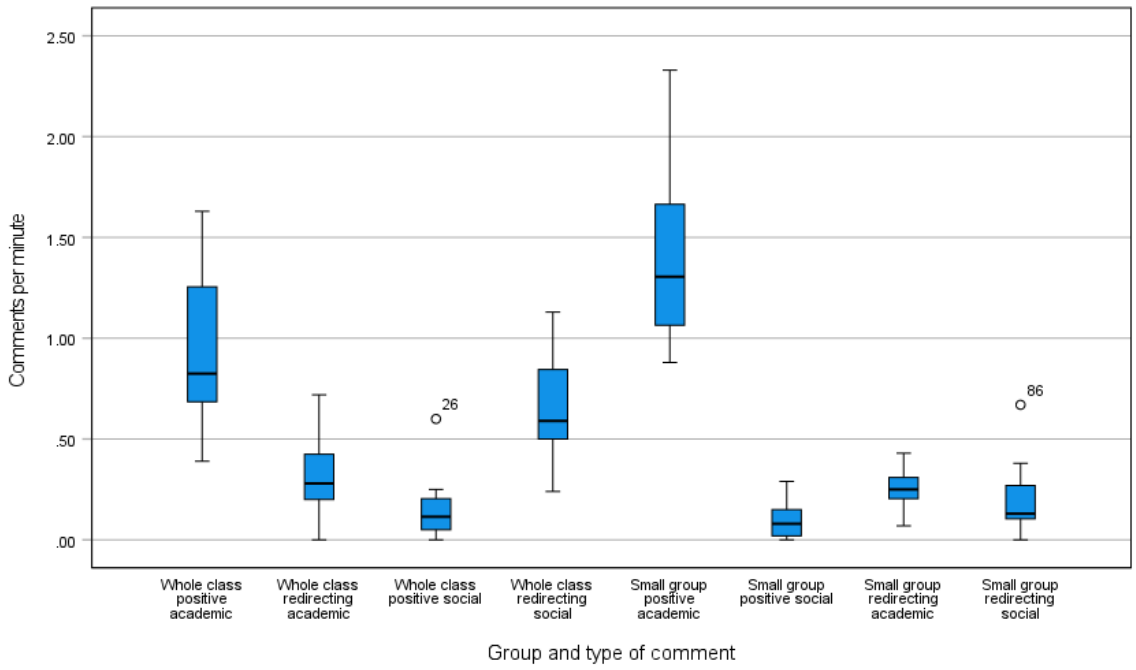
Mauchly's Test of Sphericity[a]

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
group	1.000	.000	0	.	1.000	1.000	1.000
typeofcomment	.232	14.214	5	.015	.584	.686	.333
group * typeofcomment	.236	14.050	5	.016	.526	.599	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: group + typeofcomment + group * typeofcomment

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.



Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
group	Sphericity Assumed	.019	1	.019	1.034	.331	.086
	Greenhouse-Geisser	.019	1.000	.019	1.034	.331	.086
	Huynh-Feldt	.019	1.000	.019	1.034	.331	.086
	Lower-bound	.019	1.000	.019	1.034	.331	.086
Error(group)	Sphericity Assumed	.199	11	.018			
	Greenhouse-Geisser	.199	11.000	.018			
	Huynh-Feldt	.199	11.000	.018			
	Lower-bound	.199	11.000	.018			
typeofcomment	Sphericity Assumed	15.047	3	5.016	62.655	<.001	.851
	Greenhouse-Geisser	15.047	1.752	8.586	62.655	<.001	.851
	Huynh-Feldt	15.047	2.058	7.312	62.655	<.001	.851
	Lower-bound	15.047	1.000	15.047	62.655	<.001	.851
Error(typeofcomment)	Sphericity Assumed	2.642	33	.080			
	Greenhouse-Geisser	2.642	19.277	.137			
	Huynh-Feldt	2.642	22.636	.117			
	Lower-bound	2.642	11.000	.240			
group * typeofcomment	Sphericity Assumed	2.797	3	.932	31.121	<.001	.739
	Greenhouse-Geisser	2.797	1.577	1.774	31.121	<.001	.739
	Huynh-Feldt	2.797	1.796	1.557	31.121	<.001	.739
	Lower-bound	2.797	1.000	2.797	31.121	<.001	.739
Error(group*typeofcomment)	Sphericity Assumed	.989	33	.030			
	Greenhouse-Geisser	.989	17.346	.057			
	Huynh-Feldt	.989	19.755	.050			
	Lower-bound	.989	11.000	.090			

Appendix K Individual T Test Statistics for Whole Class Compared With Small Group For All Four Types of Language

Descriptives

			Statistic	Std. Error
COMPUTE difference=WC_pos_ac - SG_pos_ac	Mean		-.4583	.08373
	95% Confidence Interval for Mean	Lower Bound	-.6426	
		Upper Bound	-.2740	
	5% Trimmed Mean		-.4537	
	Median		-.4000	
	Variance		.084	
	Std. Deviation		.29004	
	Minimum		-.93	
	Maximum		-.07	
	Range		.86	
	Interquartile Range		.52	
	Skewness		-.482	.637
	Kurtosis		-.963	1.232
Difference WC pos soc - SG pos soc	Mean		.2267	.05393
	95% Confidence Interval for Mean	Lower Bound	.1080	
		Upper Bound	.3454	
	5% Trimmed Mean		.2180	
	Median		.1400	
	Variance		.035	
	Std. Deviation		.18681	
	Minimum		.00	
	Maximum		.61	
	Range		.61	
	Interquartile Range		.31	
	Skewness		.779	.637
	Kurtosis		-.332	1.232
difference= WC_neg_ac - SG_neg_ac	Mean		-.1008	.03575
	95% Confidence Interval for Mean	Lower Bound	-.1795	
		Upper Bound	-.0222	
	5% Trimmed Mean		-.1054	

	Median		-.1200	
	Variance		.015	
	Std. Deviation		.12384	
	Minimum		-.29	
	Maximum		.17	
	Range		.46	
	Interquartile Range		.18	
	Skewness		.808	.637
	Kurtosis		.901	1.232
Difference WCneg soc - SG neg soc	Mean		.4467	.08276
	95% Confidence Interval for Mean	Lower Bound	.2645	
		Upper Bound	.6288	
	5% Trimmed Mean		.4524	
	Median		.4350	
	Variance		.082	
	Std. Deviation		.28668	
	Minimum		-.10	
	Maximum		.89	
	Range		.99	
	Interquartile Range		.45	
	Skewness		-.153	.637
	Kurtosis		-.078	1.232

Paired Samples Effect Sizes

			Standardizer [a]	Point Estimate	95% Confidence Interval	
					Lower	Upper
Pair 1	Whole class comments positive academic comments per minute - Small group comments positive academic comments per minute	Cohen's d	.29004	-1.580	-2.428	-.702
		Hedges' correction	.31189	-1.470	-2.258	-.653
Pair 2	Whole class comments positive social comments per minute - Small group comments positive social comments per minute	Cohen's d	.18598	1.214	.443	1.955
		Hedges' correction	.19999	1.129	.412	1.818
Pair 3	Whole class comments redirecting academic comments per minute - Small group comments redirecting academic comments per minute	Cohen's d	.12384	-.814	-1.459	-.143
		Hedges' correction	.13316	-.757	-1.357	-.133
Pair 4	Whole class comments redirecting social comments per minute - Small group comments redirecting social comments per minute	Cohen's d	.28615	1.555	.685	2.395
		Hedges' correction	.30770	1.446	.637	2.228

a. The denominator used in estimating the effect sizes.

Cohen's d uses the sample standard deviation of the mean difference.

Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
COMPUTE difference=WC_pos_ac - SG_pos_ac	.146	12	.200*	.933	12	.409
Difference WC pos soc - SG pos soc	.251	12	.035	.892	12	.126
difference= WC_neg_ac - SG_neg_ac	.164	12	.200*	.953	12	.686
Difference WCneg soc - SG neg soc	.158	12	.200*	.963	12	.823

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Paired Samples Test

		Paired Differences					t	df	Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			One-Sided p	Two-Sided p
Pair 1	Whole class comments positive academic comments per minute - Small group comments positive academic comments per minute	-.45833	.29004	.08373	-.64262	-.27405	-5.474	11	<.001	<.001
Pair 2	Whole class comments positive social comments per minute - Small group comments positive social comments per minute	.22583	.18598	.05369	.10766	.34400	4.206	11	<.001	.001
Pair 3	Whole class comments redirecting academic comments per minute - Small group comments redirecting academic comments per minute	-.10083	.12384	.03575	-.17952	-.02215	-2.821	11	.008	.017
Pair 4	Whole class comments redirecting social comments per minute - Small group comments redirecting social comments per minute	.44500	.28615	.08260	.26319	.62681	5.387	11	<.001	<.001

Appendix L T Test Statistics: Whole Class Positive Compared With Small Group Positive, Whole Class Redirective Compared With Small Group Redirective

Descriptives

			Statistic	Std. Error
Difference between positive whole class and positive small group	Mean		-.1163	.08633
	95% Confidence Interval for Mean	Lower Bound	-.2948	
		Upper Bound	.0623	
	5% Trimmed Mean		-.1100	
	Median		-.0350	
	Variance		.179	
	Std. Deviation		.42295	
	Minimum		-.93	
	Maximum		.61	
	Range		1.54	
	Interquartile Range		.59	
	Skewness		-.389	.472
	Kurtosis		-.571	.918

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Difference between positive whole class and positive small group	.137	24	.200*	.964	24	.517

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Comparison	WCComments - SGComments	-.11625	.42295	.08633	-.29485	.06235	-1.347	23	.096	.191

Descriptives

			Statistic	Std. Error
Difference between redirect whole class and redirect small group	Mean		.1721	.07194
	95% Confidence Interval for Mean	Lower Bound	.0233	
		Upper Bound	.3209	
	5% Trimmed Mean		.1577	
	Median		.0950	
	Variance		.124	
	Std. Deviation		.35244	
	Minimum		-.29	
	Maximum		.89	
	Range		1.18	
	Interquartile Range		.57	
	Skewness		.636	.472
	Kurtosis		-.717	.918

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Difference between redirect whole class and redirect small group	.171	24	.069	.913	24	.040

a. Lilliefors Significance Correction

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Comparison	WCComments - SGComments	.17208	.35244	.07194	.02326	.32091	2.392	23	.013	.025

Appendix M Illustrative quotations for subthemes

THEME 1 Beliefs and knowledge of pedagogy	
Linked to theories	*"That was a really nice moment for me to see because A, it was nice that she volunteered to share that, and also with just a little bit of scaffolding, she made that link to what we'd spoken about at the table before."* *"It's a bit like when they're writing, they know how to spell for example, the word 'once' when they're writing a story, coz there's so many other things to think about, they would write it as 'WUNS' because there is too much other things to think about."*
Positivity is the ideal	*"I'd like to see a shift in those [the redirecting comments] just to make them a bit more balanced in the right direction."* *"Everybody goes in with high expectations of what they're going to achieve in terms of keeping the negative language to a minimum, and the positive language to the maximum."*
Quantity of talk is important	*"Because sometimes it's, you know, it's not the right thing to interrupt the flow of teaching just to redirect a particular child's behaviour. Often, it's the best choice I think to just do it privately, quickly, you go down and then you come back."* *"But in terms of my development, it's ways of perhaps those non-verbal cues or ways that I can cut my language down perhaps."*
Clarity of expectations is important	*"I think for me as a teacher sometimes I like to reflect on how much teacher talk I'm using, and how clear my instructions are."* *"So how are they meant to know what I want if when they are doing the right thing, I'm not saying "yeah, that's the right thing, super!"* *"I want the children to know what behaviour I expect."*
Nature or structure of the curriculum	*"But today because it was new language, … new maths signs to them, um, I think it was at their level enough but me supporting them and just saying 'have a go, don't worry, then we'll talk it through and unpick it."* *"So it's kind of at the moment the focus is on building those cooperation and independent skills, so I'm trying to do,*

	design tasks at this point in the year that enable that, that then get more challenging as the lesson sort of goes on." *consolidating learning from Year three, so they had learnt it in year three, and then we were trying to just recap and consolidate I guess yeah.*
Types of language can be used strategically	*"So, I'd like to think that um, I was as positive as I could be with them to try and sort of encourage…"* *"The person sitting next to you is sitting fantastically."* *"Try not to say oh you got that wrong but rather say ok well, let's try it this way and see if we match or let's see if we can get the same answer using a different method."* *"But there's definitely a place for looking at non-examples, and especially if there's a common mistake they're making, or some kind of a common misconception, it's important to kind of address that and explain why it's wrong."*
Changing practice over time	*"When I was at school, and this is not what I want children to be like but, when I was at school, I was fearful of teachers."* *"Well, I think it's interesting just going back to the first one you showed me, which was the one from a long time ago. I do wonder whether you had less need to re-direct socially."*
THEME 2 Individual needs of children	
Cognitive needs	*"There's a particular child who is quite a bit behind the rest, even within that intervention group."* *"I choose her because she struggles with maths, but actually she was engaged during the input and the starter for five, which actually she finds quite tricky. So, the fact that she was engaged and she was able to do it was nice."*
Developmental needs which change over time	*"I think with the positive comments about social skills and behaviour, I think that is sadly something that doesn't get… I do do a few like 'well done, thank you, you can get a tribe point' and trying to do positive reinforcement, um, but with the year sixes I think that sadly we don't do enough of that, um, praising those that are doing the right thing, just because it's seen as like, they're in year six, you know. I know when I'm down in lower school I do do a bit more of that."* *"Because some of them are just coming out of year R, some of them are still working towards those early learning goals, there's still a big emphasis on those learning behaviours and their kind of, emotional side of things as well."*

Social emotional mental health needs	*"I just feel like my job for her is just to cotton onto anything that she's done really well, and just fill her up with praise and boost her confidence."* *"Because she started off writing something, and then she stopped and got distracted and wanted to talk about [sensitive subject], which I'm not gonna stop her talking about…, but she didn't do what I wanted her to do"* *"The child sat to the left of me was sort of, at the beginning of the session was just wanting that security of knowing that we were on the right track."*
THEME 3 Reflections about the self	
Language makes a positive difference	*"She also sort of struggles, but she was grasping it, but she also did need a bit of help so, it felt like I was actually helping her and teaching her."* *"We built her confidence up a little bit, and she had some praise, she then felt a little bit more enthused to go and give it a go with writing, and you know, she was in my group today, so I obviously was there if she needed the support - but actually, she had some wonderful ideas."*
Own style and personality is positive	*"…give a lot of positive praise. That's often just how I am, I like them to feel like they are in a safe space to learn and just enjoy the lesson."* *"I would say I'm a positive person. In the classroom I'm looking for children to be doing the right thing as opposed to not doing the right thing, I would hope."*
Reflecting on language	*"And I've found it really helpful, …you don't get that opportunity to have that data presented to you. It's often our own personal judgement as teachers, you know, which can clearly vary quite a lot with the reality, um, and I'd rather know what the actual reality looks like so I can go and take that and reflect on that and decide whether there's any other kind of choices that I need to make next time."* *"I could've not been thinking about that and maybe that's something I'm not conscious of that I'm doing as much perhaps."* *"I think that definitely made me reflect on my practice to maybe include a bit more positive social and behaviour comments to boost that up - and hopefully less of the redirecting discipline or social behaviour."* *"I think that's really important to see how your language is affecting how children are approaching their learning."*

THEME 4 The environment	
Physical environment	*"Often what I find, is this; things out of your control that - like a perfect example today I had children outside their classroom doing work and I had adults coming in and out and I often find the/those things out of your control will spiral."* *"We do a lot of kind of, partner talk, and there's a lot of like, always opportunity for them to kind of do some independent practice where they will need to work quietly, so that there's a calm and focused environment."*
Psychosocial environment - Interpersonal factors	*"If teachers aren't saying to children 'oh well done, great effort, I'm really proud of you, I can see you're trying really hard, I know it's tricky but you know, you're trying your best, thank you,' all of that, it's about relationships, isn't it? Teaching is all about relationships."* *"If you don't feel that your teacher values you and is proud of you and thinks you're doing a good job or trying your best, then the mindset is probably 'well why should I?'"*
Psychosocial environment - Organisational factors	*"Because it's a smaller group there's… I felt there was hardly any redirection in terms of behaviour."* *"I was a bit more direct, because when you're with a small group you don't tend to have the - I don't know, the kind of pace of the small group doesn't always lend itself to doing the silence and sitting and waiting so much."*
Psychosocial environment - School policy	*"I do feel that I do do a lot of having to pick people up … if they're not listening, or … they're not, you know, behaving in the way that our school expectations expect, and this is very much led by school culture and expectations, um. As well as my own, but I'm very aware of the expectations in school, I guess, and what we what we [sic] want from the children."* *"I think that's really important to build that confidence in those children. Um and then just reinforcing the expectations of our behaviour policy."*
Psychosocial environment - Variation across time	*"I am aware having taught through both periods [referring to the statistics from previous research in 2000 and 2019] that in 2000 I would've said I would've been doing less um looking at the behaviour and more being able to, uh, look at the academic side than I do now. So, there are behavioural changes."* *"I've found that it is that first half of the term, where obviously the learning and the task is really important, but*

	actually a lot of my input is about building in those routines and those behaviours which then will then pay off throughout the year."
THEME 5 Having clear boundaries and expectations	
Knowing what are helpful and what are unhelpful academic and learning behaviours	*"One of the children, um, made really good links back to something we taught previously."* *"That was really lovely stuff when she was able to recall that previous information really well which obviously sometimes isn't that easy for all children even when it's just been done sort of 5/10 minutes ago."* *"I think that was the exact outcome which I wanted, was to pick up the stanza and pick it up and collect that information independently, because it's a really important skill that they need for their whole class reading."* *"They weren't meeting my expectation through talking and not interacting with me in the lesson."*
Knowing what are helpful and what are unhelpful social behaviours	*"I know I made positive comment to X about giving over the pencil with the pencil grip on it for X to use."* *"I just moved him so we switch, switched places basically, so I turn my body so I can see the class better, and then he had his back to the rest of the class so he couldn't be distracted. So, I just moved him because he wasn't quite focused."* *"I would like them to be more aware of other people. So, for example when somebody's talking their sentence out loud, they should be turning round and listening to them, not having a conversation with their friend."* *"I'm trying to make them aware that if they shout out, they're not giving each other a time to think."*

Appendix N Observation and reflection checklist

These question prompts are designed for use with a supportive peer who has observed the teacher in the classroom within the previous few days.

Step 1: What might be the areas for development in my classroom? Do I have any?	Can I accurately reflect on the language I used during the observation? When did I offer positive praise? Why? When did I offer redirections? Why? Did my types of language change in different contexts? How? Do I have different reflections to my observer? Why might that be? Why might I not be sure of how much language / what type of language I used?
	Did you know? *Evidence suggests that teachers do not always realise how much their language changes between small groups and whole class teaching.* *Evidence suggests that sometimes we truly believe we are doing more of something (e.g., positive praise for good social behaviours) than we actually are.* *Two researchers suggest that teachers can be engaged in 1000 interactions a day!*
Step 2: Acknowledging where I may like to make changes.	Are there any areas where I already know I would like to improve? What are they? (e.g., certain lessons, certain age groups, certain pupils)
	Did you know? *Teachers often use higher rates of praise for younger children although they recognise that higher rates of praise are beneficial for older children too.*
Step 3: Am I motivated to make changes?	What kind of teacher am I? Do I have strong beliefs or attitudes about what kind of teacher I am? Where have these come from?
	Did you know? *One theory of the reasons underlying behaviour suggests it is a person's attitudes, motivation, and beliefs about other people's views, that influence how we behave.* *Evidence suggests that teachers have a wide variety of reasons for using the language that they do. The complexity should not be underestimated.*

Step 4: Actions and reflections to support a positive environment.	How am I enabling learning in my classroom? How am I creating a positive environment in my classroom? Do I adapt the environment (e.g. using small groups) to support in using the most helpful types and quantities of language based on my knowledge of individuals and groups? Do I adapt my language and communication based on the needs of individual pupils? Do I use lots of positive language to support a positive emotional climate? Do I use non-verbal communication to keep the lesson flowing smoothly? Do I use non-verbal communication to show I have noticed behaviours?
	Did you know? *Evidence suggests that positive environments:* • *are predictable and friendly which promotes a sense of security for pupils;* • *may be linked with overall teacher wellbeing;* • *increase pupil time on-task which may be linked to better pupil outcomes.* *Evidence suggests that teachers adapt their language according to the environment even if they are not conscious of doing so. (In small groups, positive comments about the academic work increase whilst comments to redirect social behaviours decrease.) Small groups can be a helpful part of the school day.* *There is no agreement on an 'optimum' positive to negative ratio of comments. Instead, researchers suggest monitoring and varying interactions to best support the needs of individual pupils.* • *Endeavour to increase positivity through responding to the pupil's needs rather than focusing on a rule. (Some pupils or tasks may need higher rates, some pupils or tasks may not.)* • *Adapt the environment e.g., task, timings, delivery, scaffolding, grouping.* • *Know the curriculum and who might struggle.* • *Have clear expectations and support individuals to meet these.* • *Build relationships.* *Evidence suggests effective teachers show they have noticed inappropriate behaviours but also continue with the lesson to ensure smoothness and momentum. Non-verbal communication is a powerful tool for this.*

Ajzen & Fishbein (1980), Drake & Nelson (2021), Fishbein & Azjen (1975), Fisher et al., (2015), Floress et al. (2018), Kounin (1970), McLennan et al. (2020), Reddy et al. (2013), Sabey et al., (2019), Spilt et al. (2016), Sulla et al. (2019), Swinson & Harrop (2010), Watkins & Wagner (2000).

References

Agirdag, O. (2018). The impact of school SES composition on science achievement and achievement growth: mediating role of teachers' teachability culture. *Educational Research and Evaluation*, *24*(3–5), 264–276. https://doi.org/10.1080/13803611.2018.1550838

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, *17*(4), 351–371. https://doi.org/10.1177/1094428114547952

Åhslund, I., & Boström, L. (2018). Teachers' perceptions of gender differences. *International Journal of Learning, Teaching and Educational Research*, *17*(4), 28–44.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. Prentice-Hall, Inc.

Alber, S. R., & Heward, W. L. (2000). Teaching students to recruit positive attention: A review and recommendations. *Journal of Behavioral Education*, *10*(4), 177–204. https://doi.org/10.1023/A:1012258231074

Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology*, *91*(4), 731–746. https://doi.org/10.1037/0022-0663.91.4.731

Apter, B. (2016). *A large-scale quantitative investigation of teacher-feedback and students' on-task behaviour as associated indicators of the social-emotional climate for learning in academic lessons in UK secondary schools using a systematic observation method: 'MICRO.'*

Apter, B., Arnold, C., & Swinson, J. (2010). A mass observation study of student and teacher behaviour in British primary classrooms. *Educational Psychology in Practice*, *26*(2), 151–171. https://doi.org/10.1080/02667361003768518

Apter, B., Sulla, F., & Swinson, J. (2020). A review of recent large-scale systematic UK classroom observations, method and findings, utility and impact. *Educational Psychology*, *36*(4), 367–385. https://doi.org/10.1080/02667363.2020.1802233

Babad, E. Y., Inbar, J., & Rosenthal, R. (1982). Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers. *Journal of Educational Psychology*, *74*(4), 459–474. https://doi.org/10.1037/0022-0663.74.4.459

Bailey, D. J., Bender, W. N., & Montgomery, D. L. (1983). Comparison of teacher, peer, and self-ratings of classroom and social behavior of adolescents: *Behavioural Disorders*, *8*(3), 153–160. https://doi.org/10.1177/019874298300800301

Baines, E. (2017). *Promoting effective group work in the primary classroom: A handbook for teachers and practitioners* (2nd ed.). Routledge.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Baudson, T. G., & Preckel, F. (2013). Teachers' implicit personality theories about the gifted: An experimental approach. *School Psychology Quarterly*, *28*(1), 37–46. https://doi.org/10.1037/spq0000011

Baum, J., Brill, S., Brown, J., Delpercio, A., Kahn, E., Kenney, L., & Nicoll, A. (2012). *Supporting and caring for our gender expansive youth: Lessons from the human rights campaign's youth survey*. https://www.hrc.org/supporting-and-caring-for-our-gender-expansive-youth

Beaman, R., & Wheldall, K. (2000). Teachers' use of approval and disapproval in the classroom. *Educational Psychology*, *20*(4), 431–446. https://doi.org/10.1080/713663753

Bell, J., & Waters, S. (2018). *Doing your research project* (7th ed.). McGraw-Hill Education.

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member checking: A tool to enhance trustworthiness or merely a nod to validation? *Qualitative Health Research*, *26*(13), 1802–1811. https://doi.org/10.1177/1049732316654870

Blanca, M. J., Alarcon, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29*(4), 552–557. https://doi.org/10.7334/psicothema2016.383

Blatchford, P., Baines, E., Rubie-Davies, C., Bassett, P., & Chowne, A. (2006). The effect of a new approach to group work on pupil-pupil and teacher-pupil interactions. *Journal of Educational Psychology*, *98*(4), 750–765. https://doi.org/10.1037/0022-0663.98.4.750

Boland, A., Cherry, G., & Dickson, R. (2017). *Doing a systematic review: A student's guide*. SAGE Publications Ltd.

Bond, C., Woods, K., Humphrey, N., Symes, W., & Green, L. (2013). Practitioner review: The effectiveness of solution focused brief therapy with children and families: A systematic and critical evaluation of the literature from 1990-2010. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(7), 707–723. https://doi.org/10.1111/jcpp.12058

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. London: SAGE Publications Ltd.

Braun, V., & Clarke, V. (2020). One size fits all? What counts as quality practice in (reflexive)

thematic analysis? *Qualitative Research in Psychology*, 1–25. https://doi.org/10.1080/14780887.2020.1769238

Braun, V., & Clarke, V. (2021a). Conceptual and design thinking for thematic analysis. *Qualitative Psychology*, *9*(1), 3–26. https://doi.org/10.1037/qup0000196

Braun, V., & Clarke, V. (2021b). Thematic analysis. In E. Lyons & A. Coyle (Eds.), *Analysing Qualitative Data in Psychology* (pp. 128–147).

Caldarella, P., Larsen, R. A. A., Williams, L., Downs, K. R., Wills, H. P., & Wehby, J. H. (2020). Effects of teachers' praise-to-reprimand ratios on elementary students' on-task behaviour. *Educational Psychology*, *40*(10), 1306–1322. https://doi.org/10.1080/01443410.2020.1711872

Caldarella, P., Larsen, R. A. A., Williams, L., & Wills, H. P. (2021). Effects of middle school teachers' praise-to-reprimand ratios on students' classroom behavior. *Journal of Positive Behavior Interventions*, *25*(1), 1–13. https://doi.org/10.1177/10983007211035185

Campbell, T. (2015). Stereotyped at seven? Biases in teacher judgement of pupils' ability and attainment. *Journal of Social Policy*, *44*(3), 517–547. https://doi.org/10.1017/S0047279415000227

CASP. (2019). *Critical Appraisal Skills Programme*. CASP Checklists [Online]. https://casp-uk.net/casp-tools-checklists/

Cline, T., & Ertubey, C. (1997). The impact of gender on primary teachers' evaluations of children's difficulties in school. *British Journal of Educational Psychology*, *67*(4), 447–456. https://doi.org/10.1111/J.2044-8279.1997.TB01257.X

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Collin, J., & Quigley, A. (2021). Teacher feedback to improve pupil learning. In *Education Endowment Foundation Guidance Report*. https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/feedback

Cook, C. R., Grady, E. A., Long, A. C., Renshaw, T., Codding, R. S., Fiat, A., & Larson, M. (2017). Evaluating the impact of increasing general education teachers' ratio of positive-to-negative interactions on students' classroom behavior. *Journal of Positive Behavior Interventions*, *19*(2), 67–77. https://doi.org/10.1177/1098300716679137

Coyle, A. (2021). Introduction to qualitative psychological research. In E. Lyons & A. Coyle (Eds.), *Analysing Qualitative Data in Psychology* (pp. 11–34).

Creswell, J., & Creswell, D. (2022). *Research design: Qualitative, quantitative, and mixed methods*

approaches (6th ed.). Sage Publications US.

Creswell, & Miller. (2000). Determining validity in qualitative inquiry. *Theory Into Practice*, *39*(3), 124–130. https://doi.org/10.1207/s15430421tip3903

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. SAGE Publications Ltd.

Demetriou, K. (2020). Special educational needs categorisation systems: To be labelled or not? *International Journal of Disability, Development and Education*, *69*(5), 1772–1794. https://doi.org/10.1080/1034912X.2020.1825641

Department for Education. (2023). *School teacher workforce*. School Workforce in England. https://explore-education-statistics.service.gov.uk/find-statistics/school-workforce-in-england

Department For Education. (2010). The importance of teaching. In *White Paper*. London: DfE. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/175429/CM-7980.pdf

Diamond, L. M. (2020). Gender fluidity and nonbinary gender identities among children and adolescents. *Child Development Perspectives*, *14*(2), 110–115. https://doi.org/10.1111/cdep.12366

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction*, *24*(1), 4–14. https://doi.org/10.1016/j.learninstruc.2012.06.001

Doherty, J., & Conolly, M. (1985). How accurately can primary school teachers predict the scores of their pupils in standardised tests of attainment? A study of some non-cognitive factors that influence specific judgements. *Educational Studies*, *11*(1), 41–60. https://doi.org/10.1080/0305569850110105

Doherty, J., & Hier, B. (1988). Teacher expectations and specific judgements: A small-scale study of the effects of certain non-cognitive variables on teachers' academic predictions. *Educational Review*, *40*(3), 333–348. https://doi.org/10.1080/0013191880400306

Doucet, A., & Mauthner, N. (2012). Knowing responsibly: Ethics, feminist epistemologies and methodologies. In T. Miller, M. Birch, M. Mauthner, & J. Jessop (Eds.), *Ethics in Qualitative Research* (2nd ed., pp. 122–139). SAGE Publications Ltd.

Downs, K. R., Caldarella, P., Larsen, R. A. A., Charlton, C. T., Wills, H. P., Kamps, D. M., & Wehby, J. H. (2019). Teacher praise and reprimands: The differential response of students at risk of emotional and behavioral disorders. *Journal of Positive Behavior Interventions*, *21*(3), 135–

147. https://doi.org/10.1177/1098300718800824

Drake, K. R., & Nelson, G. (2021). Natural rates of teacher praise in the classroom: A systematic review of observational studies. *Psychology in the Schools*, *58*(12), 2404–2424. https://doi.org/10.1002/pits.22602

Fagan, D. S. (2018). Addressing learner hesitancy-to-respond within initiation-response-feedback sequences. *Teaching English to Speakers of Other Languages Quarterly*, *52*(2), 425–435. https://doi.org/10.1002/tesq.421

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications Ltd.

Finch, J. (1987). The vignette technique in survey research. *Sociology*, *21*(1), 105–114.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Prentice-Hall, Inc.

Fisher, C., Berliner, D., Filby, N., Marliave, R., Cahen, L., & Dishaw, M. (2015). Teaching behaviors, academic learning time, and student achievement: An overview. *Journal of Classroom Interaction*, *50*(1), 6–24.

Fleming, N. D. (1995). I'm different; not dumb. Modes of presentation (V A R K) in the tertiary classroom. In A. Zelmer (Ed.), *Research and Development in Higher Education, Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia* (pp. 308–313).

Fleming, S. ., Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2016). Metacognition about the past and future: Quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness*, 1–12. https://doi.org/10.1093/nc/niw018

Fletcher, A. J. (2017). Applying critical realism in qualitative research: Methodology meets method. *International Journal of Social Research Methodology*, *20*(2), 181–194. https://doi.org/10.1080/13645579.2016.1144401

Flitcroft, D., & Woods, K. (2018). What does research tell high school teachers about student motivation for test performance? *Pastoral Care in Education*, *36*(2), 112–125. https://doi.org/10.1080/02643944.2018.1453858

Floress, M. T., Beaudoin, M. M., & Bernas, R. S. (2022). Exploring secondary teachers' actual and perceived praise and reprimand use. *Journal of Positive Behavior Interventions*, *24*(1), 46–57. https://doi.org/10.1177/10983007211000381

Floress, M. T., Briesch, A. M., Jenkins, L. N., & Hampton, K. A. (2021). Impact and implications statement teacher praise and reprimand: Examining the generalizability and dependability of

observational estimates. *Behavioural Disorders*, 1–11. https://doi.org/10.1177/01987429211012020

Floress, M. T., Jenkins, L. N., Reinke, W. M., & McKown, L. (2018). General education teachers' natural rates of praise: A preliminary investigation. *Behavioral Disorders*, *43*(4), 411–422. https://doi.org/10.1177/0198742917709472

Floress, M. T., Zoder-Martell, K. A., Beaudoin, M. M., & Yehling, Z. R. (2021). Teacher praise to reprimand ratios during small and large group instruction: a video pilot study. *Preventing School Failure*, *65*(3), 206–215. https://doi.org/10.1080/1045988X.2021.1898318

Fryer, T. (2022a). A critical realist approach to thematic analysis: producing causal explanations. *Journal of Critical Realism*, *21*(4), 365–384. https://doi.org/10.1080/14767430.2022.2076776

Fryer, T. (2022b). *A short guide to ontology and epistomology*. tfryer.com/ontology-guide

Geake, J. (2008). Neuromythologies in education. *Educational Research*, *50*(2), 123–133. https://doi.org/10.1080/00131880802082518

Gentrup, S., Lorenz, G., Kristen, C., & Kogan, I. (2020). Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement. *Learning and Instruction*, *66*(November 2019), 101296. https://doi.org/10.1016/j.learninstruc.2019.101296

Gentrup, S., & Rjosk, C. (2018). Pygmalion and the gender gap: do teacher expectations contribute to differences in achievement between boys and girls at the beginning of schooling? *Education Research and Evaluation*, *24*(3–5), 295–323. https://doi.org/10.1080/13803611.2018.1550840

Gentrup, S., Rjosk, C., Stanat, P., & Lorenz, G. (2018). Teachers' perceptions of students' motivation and learning behaviour and their role in biased teacher achievement expectations. *Zeitschrift Fur Erziehungswissenschaft*, *21*(4), 867–891. https://doi.org/10.1007/s11618-018-0806-2

Good, T. L. (2014). What do we know about how teachers influence student performance on standardized tests: And why do we know so little about other student outcomes? *Teachers College Record*, *116*(1), 1–41. https://doi.org/10.1177/016146811411600110

Good, T. L., & Brophy, J. E. (1994). *Contemporary educational psychology* (5th ed.). Pearson.

Good, T. L., & Brophy, J. E. (1997). *Looking in classrooms* (7th ed.). New York, NY: Longman.

Good, T. L., Sterzinger, N., & Lavigne, A. (2018). Expectation effects: Pygmalion and the initial 20 years of research. *Educational Research and Evaluation*, *24*(3–5), 99–123. https://doi.org/10.1080/13803611.2018.1548817

Gove, M. (2013). *Education reform: New national curriclulm for schools*. https://www.gov.uk/government/speeches/education-reform-new-national-curriculum-for-schools

Gülgöz, S., Edwards, D. L., & Olson, K. R. (2022). Between a boy and a girl: Measuring gender identity on a continuum. *Social Development*, *31*(3), 916–929. https://doi.org/10.1111/sode.12587

Hamilton, P. L., & Jones, L. (2016). Illuminating the "boy problem" from children's and teachers' perspectives: A pilot study. *Education 3-13 International Journal of Primary, Elementary and Early Years Education*, *44*(3), 241–254. https://doi.org/10.1080/03004279.2014.903987

Hao, L., Rubie-Davies, C. M., & Watson, P. W. S. J. (2022). Do Teachers maintain their Expectation Bias for students? A longitudinal investigation. *Social Psychology of Education*, *25*(4), 719–744. https://doi.org/10.1007/s11218-022-09714-6

Harrop, A., & Swinson, J. (2000). Natural rates of approval and disapproval in British infant, junior and secondary classrooms. *British Journal of Educational Psychology*, *70*(4), 473–483. https://doi.org/10.1348/000709900158236

Harrop, A., & Swinson, J. (2011). Comparison of teacher talk directed to boys and girls and its relationship to their behaviour in secondary and primary schools. *Educational Studies*, *37*(1), 115–125. https://doi.org/10.1080/03055691003729260

Hartas, D. (2018). Setting for English and maths: 11-year-olds' characteristics and teacher perceptions of school attitudes. *Research Papers in Education*, *33*(3), 393–410. https://doi.org/10.1080/02671522.2017.1329338

Harter, S. (2006). The self. In N. Eisenberg, R. Lerner, & W. Damon (Eds.), *Handbook of Child Psychology, social, emotional, and personality development* (pp. 505–570).

Hecht, S. A., & Greenfield, D. B. (2002). Explaining the predictive accuracy of teacher judgments of their students' reading achievement: The role of gender, classroom behavior, and emergent literacy skills in a longitudinal sample of children exposed to poverty. *Reading and Writing*, *15*(7–8), 789–809. https://doi.org/10.1023/a:1020985701556

Henderlong, J., & Lepper, M. R. (2002). The effects of praise on children's intrinsic motivation: A review and synthesis. *Psychological Bulletin*, *128*(5), 774–795. https://doi.org/10.1037/0033-2909.128.5.774

Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, *101*(3), 662–670. https://doi.org/10.1037/a0014306

Ho, A. (2004). To be labelled, or not to be labelled: that is the question. *British Journal of Learning Disabilties*, *32*, 86–92. https://doi.org/10.1038/s41587-019-0119-x

Holder, K., & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: a new look from a shifting standards perspective. *Social Psychology of Education*, *20*(3), 471–490. https://doi.org/10.1007/s11218-017-9384-z

Howe, C., Tolmie, A., Thurston, A., Topping, K., Christie, D., Livingston, K., Jessiman, E., & Donaldson, C. (2007). Group work in elementary science: Towards organisational principles for supporting pupil learning. *Learning and Instruction*, *17*(5), 549–563. https://doi.org/10.1016/j.learninstruc.2007.09.004

Jabůrek, M., Cígler, H., Valešová, T., & Portešová, Š. (2022). What is the basis of teacher judgment of student cognitive abilities and academic achievement and what affects its accuracy? *Contemporary Educational Psychology*, *69*(April). https://doi.org/10.1016/j.cedpsych.2022.102068

Jenkins, L., & Floress, M. (2015). Rates and types of teacher praise: A review and future directions. *Psychology in the Schools*, *52*(5), 463–476.

John, O. P., & Soto, C. J. (2007). The importance of being valid. In *Handbook of Research Methods in Personality Psychology* (pp. 461–494). The Guilford Press.

Jones, Jones, P., & Talbott Jones, J. L. (2007). *Tools for teaching*. F.H. Jones & Associates.

Jones, S., & Myhill, D. (2010). "Troublesome boys" and "compliant girls": gender identity and perceptions of achievement and underachievement. *British Journal of Sociology of Education*, *25*(5), 547–561. https://doi.org/10.1080/0142569042000252044

Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review*, *93*(4), 429–445. https://doi.org/10.1037//0033-295x.93.4.429

Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review*, *98*(1), 54–73. https://doi.org/10.1037/0033-295X.98.1.54

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, *9*(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3

Kaiser, J., Südkamp, A., & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology, 109*(6), 871–888.

Kelley, H. H. (1992). Common sense psychology and scientific psychology. *Annual Review of Psychology*, *43*, 1–23.

Kildea, S., Wright, J., & Davies, J. (2011). Making sense of ADHD in practice: A stakeholder review. *Clinical Child Psychology and Psychiatry*, *16*(4), 599–619. https://doi.org/10.1177/1359104510390428

Kim, Y. H., & Stormont, M. (2016). Korean early childhood educators' self-reported versus observed strategies for addressing challenging behaviors: An exploratory study. *International Journal on Disability and Human Development*, *15*(1), 43–48. https://doi.org/10.1515/ijdhd-2014-0029

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness*. Cambridge University Press.

Kounin, J. S. (1970). *Discipline and group management in classrooms*. Holt, Rinehart & Winston, Inc.

Kuklinski, M., & Weinstein, R. (2000). Classroom and grade level differences in the stability of teacher expectations and perceived differential teacher treatment. *Learning Environments Research*, *3*(1), 1–34. https://doi.org/10.1023/A:1009904718353

Lazarides, R., & Lauermann, F. (2019). Gendered paths into STEM-related and language-related careers: Girls' and boys' motivational beliefs and career plans in math and language arts. *Frontiers in Psychology*, *10*(JUN), 1–17. https://doi.org/10.3389/fpsyg.2019.01243

Lazarides, R., & Watt, H. M. G. (2015). Girls' and boys' perceived mathematics teacher beliefs, classroom learning environments and mathematical career intentions. *Contemporary Educational Psychology*, *41*, 51–61. https://doi.org/10.1016/j.cedpsych.2014.11.005

Lepper, M. (1985). Social control processes and the internalization of social values: An attributional perspective. In E. T. Higgins, D. N. Ruble, & W. W. Hartup (Eds.), *Social Cognition and Social Development*. Cambridge University Press.

Long, H. A., French, D. P., & Brooks, J. M. (2020). Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis. *Research Methods in Medicine & Health Sciences*, *1*(1), 31–42. https://doi.org/10.1177/2632084320947559

Macfarlane, B. (2009). *Researching with integrity: The ethics of academic enquiry*. Routledge.

MacQuarrie, S., Howe, C., & Boyle, J. (2012). Exploring the characteristics of small groups within science and English secondary classrooms. *Cambridge Journal of Education*, *42*(4), 527–546. https://doi.org/10.1080/0305764X.2012.733345

McEvoy, P., & Richards, D. (2006). A critical realist rationale for using a combination of quantitative and qualitative methods. *Journal of Research in Nursing*, *11*(1), 66–78.

https://doi.org/10.1177/1744987106060192

McLennan, J. D., Sampasa-Kanyinga, H., Georgiades, K., & Duku, E. (2020). Variation in teachers' reported use of classroom management and behavioral health strategies by grade level. *School Mental Health*, *12*(1), 67–76. https://doi.org/10.1007/s12310-019-09341-1

Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, *65*, 48–60. https://doi.org/10.1016/j.tate.2017.02.021

Merrett, F., & Wheldall, K. (1986). Observing pupils and teachers in classrooms (OPTIC): A behavioural observation schedule for use in schools. *Educational Psychology*, *6*(1), 57–70. https://doi.org/10.1080/0144341860060107

Monro, S. (2005). Beyond male and female: Poststructuralism and the spectrum of gender. *International Journal of Transgenderism*, *8*(1), 3–22. https://doi.org/10.1300/J485V08N01_02

Nafpaktitis, M., Mayer, G. R., & Butterworth, T. (1985). Natural rates of teacher approval and disapproval and their relation to student behavior in intermediate school classrooms. *Journal of Educational Psychology*, *77*(3), 362–367. https://doi.org/10.1037//0022-0663.77.3.362

Newman, D., Griffin, P., & Cole, M. (1989). *The construction zone: Working for cognitive change in school*. Cambridge University Press.

Newton, P. M. (2015). The learning styles myth is thriving in higher education. *Frontiers in Psychology*, *6*(DEC), 1–5. https://doi.org/10.3389/fpsyg.2015.01908

Organisation for Economic Co-operation and Development. (2016). *PISA 2015 results (Volume I) Excellence and equity in education.* OECD. https://doi.org/10.1787/9789264266490-EN

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*. https://doi.org/10.1136/BMJ.N71

Pirskanen, H., Jokinen, K., Karhinen-Soppi, A., Notko, M., Lämsä, T., Otani, M., Meil, G., Romero-Balsas, P., & Rogero-García, J. (2019). Children's emotions in educational settings: Teacher perceptions from Australia, China, Finland, Japan and Spain. *Early Childhood Education Journal*, *47*(4), 417–426. https://doi.org/10.1007/S10643-019-00944-6/TABLES/1

Plewis, I. (1997). Inferences about teacher expectations from national assessment at key stage

one. *British Journal of Educational Psychology*, *67*(2), 235–247. https://doi.org/10.1111/J.2044-8279.1997.TB01240.X

Prochaska, J. O., & Di Clemente, C. C. (1982). Transtheoretical therapy: Toward a more integrative model of change. *Psychotherapy*, *19*(3), 276–288. https://doi.org/10.1037/h0088437

Ready, D. D., & Chu, E. M. (2015). Sociodemographic inequality in early literacy development: The role of teacher perceptual accuracy. *Early Education and Development*, *26*(7), 970–987. https://doi.org/10.1080/10409289.2015.1004516

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, *48*(2), 335–360. https://doi.org/10.3102/0002831210374874/FORMAT/EPUB

Reay, D. (2022). The slide to authoritarianism in English schools. *Forum*, *64*(3), 126–139.

Reddy, L. A., Fabiano, G. A., Dudek, C. M., & Hsu, L. (2013). Instructional and behavior management practices implemented by elementary general education teachers. *Journal of School Psychology*, *51*(6), 683–700. https://doi.org/10.1016/j.jsp.2013.10.001

Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom-level positive behavior supports in schools implementing sw-pbis: Identifying areas for enhancement. *Journal of Positive Behavior Interventions*, *15*(1), 39–50. https://doi.org/10.1177/1098300712459079

Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review*, *37*(3), 315–332. https://doi.org/10.1080/02796015.2008.12087879

Riegle-Crumb, C., & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender and Society*, *26*(2), 290–322. https://doi.org/10.1177/0891243211434614

Riener, C., & Willingham, D. (2010). The myth of learning styles. *Change: The Magazine of Higher Learning*, *42*(5), 32–35. https://doi.org/10.1080/00091383.2010.503139

Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental Psychology*, *50*(4), 1262–1281. https://doi.org/10.1037/a0035073

Roer-strier, D., & Kurman, J. (2009). Combining qualitative and quantitative methods to study perceptions of immigrant youth. *Journal of Cross-Cultural Psychology*, *40*(6), 988–995.

Rogers, C. R. (1959). *A theory of therapy, personality, and interpersonal relationships, as*

developed in the client-centred framework. McGraw-Hill.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. *The Urban Review*, *3*(September), 16–20.

Royer, D. J., Lane, K. L., Dunlap, K. D., & Ennis, R. P. (2019). A Systematic Review of Teacher-Delivered Behavior-Specific Praise on K–12 Student Performance. *Remedial and Special Education*, *40*(2), 112–128. https://doi.org/10.1177/0741932517751054/FORMAT/EPUB

Rubie-Davies, C. M., & Peterson, E. R. (2016). Relations between teachers' achievement, over- and underestimation, and students' beliefs for Māori and Pākehā students. *Contemporary Educational Psychology*, *47*, 72–83. https://doi.org/10.1016/j.cedpsych.2016.01.001

Rubie-Davies, C. M., Watson, P. W. S. J., Flint, A., Garrett, L., & McDonald, L. (2018). Viewing students consistently: how stable are teachers' expectations? *Educational Research and Evaluation*, *24*(3–5), 221–240. https://doi.org/10.1080/13803611.2018.1550836

Rubie-Davies, C. M., Weinstein, R. S., Huang, F. L., Gregory, A., Cowan, P. A., & Cowan, C. P. (2014). Successive teacher expectation effects across the early school years. *Journal of Applied Developmental Psychology*, *35*(3), 181–191. https://doi.org/10.1016/j.appdev.2014.03.006

Sabey, C. V., Charlton, C., & Charlton, S. R. (2019). The "magic" positive-to-negative interaction ratio: Benefits, applications, cautions, and recommendations. *Journal of Emotional and Behavioral Disorders*, *27*(3), 154–164. https://doi.org/10.1177/1063426618763106

Sazak-Pinar, E., & Guner-Yildiz, N. (2013). Investigating teachers' approval and disapproval behaviors towards academic and social behaviors of students with and without special needs. *Educational Sciences: Theory and Practice*, *13*(1), 551–556. www.edam.com.tr/estp

Schwappach, D. L. B., Frank, O., & Davis, R. E. (2013). A vignette study to examine health care professionals' attitudes towards patient involvement in error prevention. *Journal of Evaluation in Clinical Practice*, *19*(5), 840–848. https://doi.org/10.1111/j.1365-2753.2012.01861.x

Sinclair, & Coulthard, R. (1975). *Toward an analysis of discourse*. Oxford University Press.

Sinclair, S., Huntsinger, J., Skorinko, J., & Hardin, C. D. (2005). Social tuning of the self: Consequences for the self-evaluations of stereotype targets. *Journal of Personality and Social Psychology*, *89*(2), 160–175. https://doi.org/10.1037/0022-3514.89.2.160

Spilt, J. L., Leflot, G., Onghena, P., & Colpin, H. (2016). Use of praise and reprimands as critical ingredients of teacher behavior management: effects on children's development in the context of a teacher-mediated classroom intervention. *Prevention Science*, *17*(6), 732–742.

https://doi.org/10.1007/s11121-016-0667-y

Stichter, J. P., Lewis, T. J., Whittaker, T. A., Richter, M., Johnson, N. W., & Trussell, R. P. (2009). Assessing teacher use of opportunities to respond and effective classroom management strategies: Comparisons among high- and low-risk elementary schools. *Journal of Positive Behavior Interventions*, *11*(2), 68–81. https://doi.org/10.1177/1098300708326597

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762. https://doi.org/10.1037/a0027627

Sulla, F., Armenia, E., & Rollo, D. (2019). Natural rates of teachers' approval and disapproval in italian primary and secondary schools classroom. In G. Minati, M. Abram, & E. Pessa (Eds.), *Systemics of Incompleteness and Quasi-Systems* (pp. 335–342). Springer, Cham. https://doi.org/10.1007/978-3-030-15277-2_27

Swinson, J., & Harrop, A. (2010). An examination of the effects of a short course aimed at enabling teachers in infant, junior and secondary schools to alter the verbal feedback given to their pupils. *Educational Studies*, *31*(2), 115–129. https://doi.org/10.1080/03055690500095316

Tandler, N., & Dalbert, C. (2020). Always look on the bright side of students: does valence of teacher perceptions relate to students' educational performance? *Social Psychology of Education*, *23*(5), 1121–1147. https://doi.org/10.1007/S11218-020-09573-Z

Thompson, B. (2011). Reassessing gender issues in the primary classroom. In G. Knowles (Ed.), *Supporting Inclusive Practice* (2nd ed., pp. 46–62). Oxon: Routledge.

Tiedemann, J. (2000). Gender-related beliefs of teachers in elementary school mathematics. *Educational Studies in Mathematics*, *41*, 191–207. https://link.springer.com/content/pdf/10.1023/A:1003953801526.pdf

Timmermans, A. C., de Boer, H., & van der Werf, M. P. C. (2016). An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes. *Social Psychology of Education*, *19*(2), 217–240. https://doi.org/10.1007/s11218-015-9326-6

Timmermans, A. C., Rubie-Davies, C. M., & Rjosk, C. (2018). Pygmalion's 50th anniversary: the state of the art in teacher expectation research. *Educational Research and Evaluation*, *24*(3–5), 91–98. https://doi.org/10.1080/13803611.2018.1548785

Timmermans, A. C., Rubie-Davies, C. M., & Wang, S. (2021). Adjusting expectations or maintaining first impressions? The stability of teachers' expectations of students' mathematics achievement. *Learning and Instruction*, *75*, 101483. https://doi.org/10.1016/j.learninstruc.2021.101483

Trouilloud, D. O., Sarrazin, P. G., Martinek, T. J., & Guillet, E. (2002). The influence of teacher expectations on student achievement in physical education classes: Pygmalion revisited. *European Journal of Social Psychology*, *32*(5), 591–607. https://doi.org/10.1002/ejsp.109

Wang, S., Meissel, K., & Rubie-Davies, C. M. (2021). Teacher expectation effects in Chinese junior high schools: Exploring links between teacher expectations and student achievement using a hierarchical linear modelling approach. *Social Psychology of Education*, *24*(5), 1305–1333. https://doi.org/10.1007/S11218-021-09654-7

Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, *24*(3–5), 124–179. https://doi.org/10.1080/13803611.2018.1548798

Waring, H. Z. (2008). Using explicit positive assessment in the language classroom: IRF, feedback, and learning opportunities. *Modern Language Journal*, *92*(4), 577–594. https://doi.org/10.1111/j.1540-4781.2008.00788.x

Watkins, C., & Wagner, P. (2000). *Improving School Behaviour*. Paul Chapman Publishing.

Woods, K. (2020). *Critical appraisal frameworks: Quantitative research framework*. Manchester:The University of Manchester (Education and Psychology Research Group).

Zajda, J. (2021). *Globalisation and education reforms: Creating effective learning environments*. Springer Nature Switzerland. https://doi.org/10.1007/978-94-024-1204-8_1

Zare-ee, A., & Hejazi, S. Y. (2019). The effects of EFL teachers' gender and experience on interaction patterns in undergraduate content classes. *English Teaching and Learning*, *43*(3), 255–275. https://doi.org/10.1007/s42321-019-00028-z

Zeichner, K., & Liston, D. (1987). Teaching student teachers to reflect. *Harvard Educational Review*, *57*(1), 23–49. https://doi.org/10.17763/haer.57.1.j18v7162275t1w3w