Voice Symptoms and Wellbeing in Teachers Working in England

Emily Sharp

In partial fulfilment for the degree of Masters of Philosophy

Birmingham City University. Department of Health, Education and Life Sciences

Contents

Figures and Tables

Acknowledgments

The completion of my MPhil was made possible by many people. Thanks in particular go to all the teachers and schools that gave up their valuable time to participate in my research and provide such invaluable data, without whom there would be no MPhil. I would also like to thank Dr Helen Jenkins and colleagues in the Speech and Language Therapy Department at BCU who have supported me throughout my MPhil journey.

Sincere thanks to my supervisors, Dr Kate Thompson and Dr Robert Cook who always believed in this project and gave me such helpful feedback and constructive advice during the writing up of the thesis. They helped give me the confidence to keep going, during what was at times a difficult process. Particular thanks also go to Dr Robert Cook and Dr Christina Easter for their statistical help and guidance, and willingness to answer my many questions, making the data analysis more enjoyable and less arduous than I had thought it would be. Also, many thanks to Dr Philip Dee who read my thesis and gave me valuable feedback.

Huge thanks go to all my friends and colleagues who were always there for moral support: Laura Maguire, Claire Thompson, Hannah Harvey, Ana Barrios, Sarah Bett, Maria Zachariades, Roisin Bailey and Tamar Shiboleth.

Finally, thanks go to Dr Alex Kendal and the generosity of the Health, Education and Life Sciences Doctoral College at BCU, who funded and supported my MPhil.

Abstract

Background

The literature indicates that teachers are at greater risk of getting voice problems than the general population. In order to help prevent and treat voice problems, research has been undertaken to identify risk factors. Negative psychological factors such as depression and common mental disorders have been found to be associated with voice problems in teachers. However, there is little research with teachers that investigates the relationship between positive psychological factors such as wellbeing and voice problems. Although negative and positive mental states are on a continuum of psychological health, they are measuring separate constructs and thus need to be investigated separately.

Aims

The primary objective of this study was to explore the relationship between voice symptoms and wellbeing in teachers working in primary and secondary schools in England. The study also examined the association between other risk factors and voice symptoms.

Methods

A cross sectional study was conducted between November 2017 and February 2018 using web-based self-administered questionnaires to collect data. All schools in England, including independent fee-paying schools, were invited to participate. Information was obtained on symptoms, wellbeing, health, lifestyle, sociodemographic factors and environmental factors. Voice symptoms were measured using The Voice Symptom Scale (VoiSS) and wellbeing was measured using the Warwick Edinburgh Mental Wellbeing Scale (WEMWBS). Analysis was conducted using a linear multi-level regression model.

Results

A total of 1205 teachers from 608 schools participated. Participants were primarily female (80%), white (93%) with a mean age of 39. The mean score on the VoiSS was 23 and the median was 20. A statistically significant negative relationship between voice symptoms and wellbeing was identified (-0.31 95% CI -0.41, -0.20 P=<0.001). Other factors found to be statistically significantly associated with voice symptoms were age (0.10 95% CI 0.02, 0.18 P= 0.015) and the likelihood of having gastroesophageal reflux (1.29 95% CI 0.87, 1.70 P=<0.001). VoiSS scores were significantly lower for male teachers (-3.48 95%CI -5.59,-1.37 P=0.001), for teachers who never spoke over background

noise (-5.39 95%CI -9.19, -1.61 P=0.005), teachers who never spoke louder than normal (-11.35 95% CI -15.73, -6.98 P=<0.001) or sometimes spoke louder than normal (-8.23 95%CI -11.26, -5.20 P=<0.001) and teachers who taught in smaller class sizes (-2.21 95% CI -3.99, -0.43 P=0.015). Teachers with a respiratory infection not confirmed by a doctor had significantly higher VoiSS scores compared to those with no respiratory infection (6.23 95% CI 3.76, 8.77 P=<0.001), whereas teachers who had a respiratory infection confirmed by a doctor had lower VoiSS scores compared to those with no respiratory infection (-4.15 95% CI -5.91, -2.39 P=<0.001). Number of years teaching, hours teaching per week, deprivation of school, voice training, teaching subject, smoking and asthma were not associated with voice symptoms.

Conclusions

This study suggests that there is an association between vocal symptoms and wellbeing in school teachers in England. These findings indicate that teachers vocal functioning may be improved by enhancing their wellbeing. Thus, schools implementing strategies focussed on wellbeing may be beneficial.

Chapter 1. Introduction

1.0 Introduction

Voice problems within the teaching profession are a cause for concern (National Education Union, 2019a) . Teachers have been identified as a group particularly at risk for voice problems with prevalence rates reported ranging from 8% to 81% (Alva et al., 2017; Trinite, 2017). The impact of voice problems among teachers is significant, creating substantial financial costs (National Education Union, 2019) and may be detrimental to students' learning and achievement (Chui and Ma, 2018; Lyberg-Åhlander et al., 2015b). To address these issues, it is important to know what the risk factors for voice problems are, so preventative measures can be put in place. There have been many studies with teachers investigating risk factors, but due to their methodological differences, results can be difficult to compare. Additionally, methodological limitations in the studies, such as statistical design and inadequate reporting mean that the current evidence for the associations between risk factors and voice problems is not robust.

Wellbeing is an important factor to consider that may be associated with voice problems. Governments and occupational organisations are increasingly interested in measuring positive mental health. For example, Health Survey for England, which is instrumental in helping inform health policy, includes the Warwick Edinburgh Mental WellBeing Scale (WEMWBS) in its survey. Educational organisations such as the teachers' union, NASUWT, and the charity, Education Support, support teachers' wellbeing and work with organisations to highlight the importance of mental health and wellbeing in teachers (Education Support, 2020; NASUWT, 2020). However, no literature has been identified that investigates the association of wellbeing with voice problems in teachers. Currently the literature suggests that negative psychological states such as depression and anxiety are predictive of voice problems (da Rocha et al., 2015; de Medeiros et al., 2008), but this does not mean wellbeing will be associated. Low wellbeing does not necessarily mean that someone has a psychological problem, and conversely, the absence of negative psychological states does not mean someone has high wellbeing (Huppert and Whittington, 2003; Trudel-Fitzgerald et al., 2019). Although well-being is on a continuum of psychological health, it is a separate construct to negative states such as depression and anxiety, and thus needs to be measured separately.

This study will be the first study to investigate the association of wellbeing and voice problems in teachers and will be the first nationwide study in England to look at risk factors for voice problems with teachers. A recent study has been undertaken in the North West of England with 454 teachers (Gadepalli et al., 2019) but does not investigate wellbeing. In order to carry out the current study, a

cross-sectional design was employed with the target population being teachers working in England. This chapter provides a further discussion around the prevalence of voice problems and their impact, and an initial discussion around the evidence base for wellbeing and other risk factors associated with voice problems. It provides a background and a definition of voice symptoms, the philosophical approach to the study, the study's rationale, its aims and objectives, and an overview of the thesis.

1.1 Voice Symptoms

The terms voice problem, voice symptoms, voice disorder and dysphonia are often used interchangeably in the literature, although definitions do point to differences. Dysphonia and voice disorder imply clinical assessment. Indeed dysphonia has been defined as "altered voice quality, pitch, loudness or vocal effort that impairs communication as assessed by a clinician and/ or reduces quality of life" Stachler et al. (2018, 1). A voice disorder has been defined as a change in the quality, pitch and loudness of the voice that is inappropriate for the person's age, gender and culture (Aronson and Bless, 2009; Mathieson and Greene, 2001). The term voice problem is generally used in research to refer to people who report voice difficulties regardless of whether they have been assessed by a clinician. One of the difficulties in assessing the literature is there is no standardised definition for a voice problem and no gold standard for assessment. Therefore, the definitions that are used and tools of measurement often vary between studies, which can make comparisons difficult. Voice symptoms refer to abnormal vocal characteristics, including emotional difficulties that make up a voice disorder or voice problem, and is what this study will be measuring.

In order to measure vocal symptoms and voice problems, methods and tools have been developed to allow an individual to assess their own voice. These include patient reported outcome measures (PROMS) that allow individuals to rate their voice symptoms, including physical symptoms, day-to-day functioning, and the emotional impact of their voice problem. Psychometrically tested tools such as the Voice Handicap Index (VHI) (Jacobson et al., 1997), the Voice Symptom Scale (VoiSS) (Deary et al., 2003), and the Vocal Performance Questionnaire (VPQ) (Carding et al., 1999) have been developed that are used in clinical practice and in research. In the literature regarding teachers and risk factors different ways of measuring voice problems, ranging from single questions to validated clinical tools have been undertaken. These will be critically appraised in the literature review.

1.2 Prevalence of voice problems in teachers

The literature suggests that there is a substantially higher prevalence of voice disorders among teachers than the general population (Behlau et al., 2012; Roy et al., 2004; Seifpanahi et al., 2016;

Sliwinska-Kowalska et al., 2006; Smith et al., 1998b; Van Houtte et al., 2011). Prevalence refers to the proportion of a population that has a condition at a particular point in time (Webb and Bain, 2011). A study undertaken by The Association of Teachers and Lecturers (2008) identified that 68% of teachers in primary schools in the UK had experienced voice problems that they felt had been caused by their job. Furthermore, a recent study carried out with teachers in the North West of England reported a prevalence of 30% for teachers compared with 9% for non-teachers (Gadepalli et al., 2019).

Table 1 shows the differences in prevalence figures for voice problems between teachers and non-teachers recorded in the literature. In all cases, prevalence was greater for teachers than non-teachers. As demonstrated, there is a wide range of prevalence figures between studies. This may be due to methodological differences, such as different ways of measuring voice problems or unique social and cultural factors. Different types of prevalence have been collected, point prevalence (current prevalence), 12-month prevalence, career and lifetime prevalence. Figures for lifetime and career prevalence need to be interpreted with caution as recall may not be accurate when considering events over a long period of time (Althubaiti, 2016; Sudman and Bradburn, 1973). Despite this, figures suggest that teachers are at greater risk of getting voice problems than non-teachers and therefore research with this population needs to be considered so measures can be introduced to help reduce prevalence.

Table 1: Prevalence Figures for Voice Problems in Teachers and Non-Teachers

Author	Country	*Point prev teachers	Point prev non-teachers	**12 month prev teachers	12 month prev non-teachers	***Career prev teachers	Career prev non-teachers	****Life time prev teachers	Life time prev non-teachers
Smith et al 1998	USA							32%	1%
Roy et al 2004	USA	11%	6.2%					57.7%	28.8%
Sliwinska-Kowalska et al 2006	Poland	68.2%	32.5%					68.7%	36%
van Houtte et al 2011	Belgium					51.2%	27.4%		
Behlau et al 2012	Brazil	11.6%	7.5%					63%	35.8%
Seifpanahi et al 2015	Iran	33.6%	23%	44.8%	28.2%	54.6%	21.1%	61.5%	31.7%
Gadepalli et al	England			30%	9%				

*Point prev refers to the prevalence rates of current voice problems
** 12-month prev refers to the prevalence rate of voice problems in the last 12 months
*** Career prev refers to the prevalence rate of voice problems over a teacher's teaching career

**** Lifetime prev refers to the prevalence rate of voice problems over a teacher's life.

1.3 Impact of voice problems

Prevalence rates themselves are not sufficient to warrant further research; the impact of the problem needs to be assessed. Voice problems pose difficulties in several areas. As well as the personal impact on teachers themselves, there are financial implications for schools due to staff absence and negative impacts on students' learning. The estimated cost of voice problems is high. A survey undertaken by the Royal National Institute for the Deaf in 2008 suggested more than 70,000 teaching days a year were lost in the UK because of voice problems, which they estimated to cost about £15 million a year (National Education Union, 2019). Studies have compared teachers and non-teachers in terms of absenteeism, ability to work effectively, and the consideration of changing professions due to a voice problem (Behlau et al., 2012; de Jong et al., 2006; Roy et al., 2004; Smith et al., 1998b; Van Houtte et al., 2011). In all cases, the teachers were significantly more likely than non-teachers to be absent from work due to their voice, were more likely to say that their voice problems prevented them from working effectively, and were more likely to consider changing their occupation in the future due to a voice problem. These studies all suggest that the impact on teachers and schools including the financial burden of voice problems is likely to be higher in teaching than many other professions.

Voice problems may influence a student's learning, due not only to teachers being absent, but may hinder a child's ability to understand the information they are being taught, and thus negatively affect educational achievement. Studies with teachers have found that primary school children perform more poorly on language comprehension tasks, even with mildly dysphonic voices than when hearing a 'normal' voice (Chui and Ma, 2018; Lyberg-Åhlander et al., 2015a; Rogerson and Dodd, 2005). A recent study by Schiller et al. (2020) found that 5 and 6 year olds listening to a voice with impaired quality had significantly lowered performance and increased response time on a speech perception task and lowered performance on a listening comprehension task. This suggests that having a teacher with a voice problem, even if mild is likely to have a negative effect on children's learning and subsequent achievement.

The literature also indicates that voice problems can negatively affect an individual's life. A study conducted by Smith et al. (1996) suggested that those with voice disorders had reduced quality of life compared to controls, including more adversely affected social interactions, more limited participation in social activities, lower professional self-esteem, and were more likely to report depression than controls. Other studies suggest those with voice disorders have a lower health-related quality of life than those without voice disorders (Benninger et al., 1998; Krischke et al., 2005; Wilson et al., 2002). As teachers rely on their voices as an important communication tool

(Martin and Darnley, 2004) their quality of life may be especially affected. Indeed Aghadoost et al. (2016) found that female teachers with voice problems had significantly poorer quality of life than those without voice problems. Therefore, due to the negative impact voice problems can have in teachers, research with this population needs to be considered.

1.4 Risk Factors for Voice Problems

Researching risk factors for voice problems is important as it has the potential to inform the clinical evidence base for both the prevention and treatment of voice problems. Having a clearer idea of the risk factors associated with voice problems can influence treatment given to patients by clinicians working in the field of voice, and school policies.

The risk factor of primary interest in this study is wellbeing and it will investigate whether there is an association between wellbeing and voice symptoms. Other risk factors and their evidence base will also be discussed.

1.4.1 Wellbeing

The World Health Organisation (WHO) defines wellbeing as a state where "the individual realises his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully and is able to make a contribution to his or her community" (World Health Organisation, 2001). The importance of mental wellbeing is increasingly being recognised as essential for the overall functioning and physical health (World Health Organisation, 2013). The increase in mental health awareness generally, has not only been improved by research, but media campaigns with charities such as Heads Together have helped open up the conversation around mental health, including within schools (Heads Together, 2018). Mental health in schools has been identified as an important issue. A study by Kidger et al. (2015) found that teachers in England had lower wellbeing than the general adult population. Interventions such as the Wellbeing in Secondary Education (WISE) Project have been set up to help improve mental health support in schools (Evans et al., 2018). Therefore, further research involving the wellbeing of teachers in the UK would seem to be valuable.

Although there is a wealth of research looking at the associations of mental disorders and voice problems studies focusing on wellbeing are scarce. One study was identified that looked at factors associated with wellbeing in elementary teachers in Brazil (Molina et al., 2017). It found a significant association with those who were classified as 'champions of abuse' on the Vocal Behaviour Profile (Behlau et al., 2018), indicating those that often or constantly misuse their voices have lower wellbeing. As vocal abuse is linked to voice problems, it may therefore be possible that voice problems are linked to wellbeing. However, the Vocal Behaviour Profile does not directly measure voice problems, so results need to be extrapolated with caution.

The research investigating negative mental states and voice problems includes case control studies that suggest patients with voice disorders are more likely to have anxiety and depression than healthy controls. For example, Willinger et al. (2005) found that patients with functional voice disorders had significantly higher depressive symptoms and symptoms of anxiety than healthy controls. Deary et al. (1997) suggested that men and women diagnosed with dysphonia had significantly greater scores on the General Health Questionnaire (GHQ) (Goldberg and Williams, 1988), a tool to screen for mental health problems, than healthy controls, indicating higher levels of mental health problems. It must be noted that these studies only include participants who have sought treatment for voice problems, so may not be generalisable to people with voice problems who do not seek treatment. Risk factor studies with teachers however, have also suggested a significant association between negative psychological states and having voice problems including mental disorders (da Rocha et al., 2017; da Rocha and Souza, 2013; de Medeiros et al., 2008), depression (da Rocha et al., 2015; Nerriere et al., 2009) and anxiety (Moy et al., 2015; Nerriere et al., 2009). These will be discussed further in the literature review.

Although poor mental health and positive mental health are linked, for example, poor wellbeing has been shown to be predictive of poor mental health (Keyes et al., 2010; Winefield et al., 2012), wellbeing is not merely the absence of mental health problems (Trudel-Fitzgerald et al., 2019). Although wellbeing and psychological problems are moderately correlated (Huppert and Whittington, 2003; Tennant et al., 2007) they do measure different constructs. Huppert and Whittington (2003) found some independence between negative and positive mental health. They found many participants scoring low scores on both positive and negative scales, and individuals in the top quartile of the positive measure reported symptoms of mental health problems.

Arguably, measuring wellbeing rather than psychological problems, will be more helpful to schools. Currently, the language around occupational health focusses on mental health and wellbeing and therefore a measure of wellbeing rather than measures of anxiety and depression may be more meaningful for schools. Policies centre on how to improve wellbeing, and therefore if associations are found between wellbeing and voice problems, then schools may be more likely to implement strategies. Therefore, due to the paucity in research looking at the association between wellbeing and voice disorders this study would be a useful addition to the literature.

1.4.2 Evidence Base for Other Risk Factors

It is not enough to look only at the association between voice problems and wellbeing in a model. Other factors that may be confounding need to be included in the model. For example, factors such as gender, age and health conditions. Furthermore, including other factors will help provide further

information to the clinical evidence base. There has been a range of risk factors that have been investigated for their association to voice problems. These can be categorised into environmental, behavioural, health, lifestyle, psychological and sociodemographic factors. Table 2 shows the extent of the risk factors with teachers and other populations that have been researched empirically and show a possible link with voice problems. Due to the magnitude of risk factors investigated, it is not possible to discuss every factor that has been researched, in this paper. The focus of the investigation will be on assessing the evidence base of risk factors used in research with teachers. This will be discussed further in the literature review.

Table 2: Risk Factors

Category	Possible Risk Factors
Environmental Factors	• Dry air (Cutiva and Burdorf, 2016; Levendoski et al., 2014) • Poor ventilation (de Medeiros et al., 2008) • Noise (Phadke et al., 2019; Portela et al., 2018; Rossi-Barbosa et al., 2016; Sampaio et al., 2012) • Poor acoustics (Cutiva and Burdorf, 2014) • Temperature changes (van Houtte et al., 2012)
Behavioural Factors	• Speaking loudly/shouting (Byeon, 2019; Chen et al., 2010; Fontan et al., 2016; Ubillos et al., 2015) • Speaking against background noise (Kyriakou et al., 2018; Lee et al., 2010) • Excessive voice use (de Alvear et al., 2011; Rossi-Barbosa et al., 2016) • Microphone use (Charn and Mok, 2012; Sampaio et al., 2012) • Throat clearing (Devadas et al., 2017b; Devadas et al., 2015; Kyriakou et al., 2018)
Health	• Hypertension (Bainbridge et al., 2017) • Tinnitus (Bainbridge et al., 2017) • Upper airway problems/infections (Akinbode et al., 2014; Byeon, 2019; Devadas et al., 2017a) • Laryngitis (Lee et al., 2010) • Sleep problems (Ferreira et al., 2010; Rocha and Behlau, 2018; Roy et al., 2019) • Reflux/heartburn (Alanazi et al., 2018; Charn and Mok, 2012; Lechien et al., 2020; Sampaio et al., 2012) • Thyroid problems (Charn and Mok, 2012; Devadas et al., 2017a) • Rhinitis/ allergies (Charn and Mok, 2012; Devadas et al., 2015; Sampaio et al., 2012) • Asthma (Devadas et al., 2015; Dogan et al., 2007; Lee et al., 2010) • Rheumatoid Arthritis/autoimmune disease (Liu et al., 2012; Roy et al., 2016; Speyer et al., 2008) • Sjögren's Syndrome (Kim et al., 2018; Pierce et al., 2015; Saltürk et al., 2017) • Hearing impairment (Cutiva and Burdorf, 2014) • Fibromyalgia (Gurbuzler et al., 2013) • Third trimester of pregnancy (Kosztyła-Hojna et al., 2018; Saltürk et al., 2016) • Chemoradiotherapy (Lazarus, 2009) • Inhaled corticosteroids (Erickson and Sivasankar, 2010; Pinto et al., 2013; Spantideas et al., 2017) • Cough (Gadepalli et al., 2019; Kyriakou et al., 2018)
Lifestyle	• Dehydration (Cannes do Nascimento et al., 2020; Miri et al., 2012; Verdolini et al., 2002) • Caffeine consumption (Akinbode et al., 2014; Byeon, 2019) • Smoking cigarettes (Byeon and Cha, 2020; Byeon and Lee, 2013; Tafiadis et al., 2017) • Alcohol consumption (Byeon, 2016; Rossi-Barbosa et al., 2016; Spantideas et al., 2015)
Psychological	• Anxiety (Bainbridge et al., 2017; Moy et al., 2015; Siupsinskiene et al., 2011) • Stress (Devadas et al., 2017a; Fellman and Simberg, 2017; Gassull et al., 2010; Kyriakou et al., 2018) • Mental Disorder (da Rocha et al., 2017; da Rocha and Souza, 2013; de Medeiros et al., 2008) • Psychological distress (van Houtte et al., 2012) • Depression (da Rocha et al., 2015; Marmor et al., 2016)
Socio-demographic	• Female (Albustan et al., 2018; Bainbridge et al., 2017; Byeon, 2019; Lyberg-Åhlander et al., 2019) • Age (Fontan et al., 2016; Lia et al., 2019; Lyberg-Åhlander et al., 2019; Reed and Sims, 2016), • Ethnicity (Reed and Sims, 2016), • Education level (Kim et al., 2015)

1.5 Philosophical Approach.

The philosophical approach that underlies this research is pragmatism. The foundations of the approach are commonly associated with the American philosophers Charles Sanders Peirce, William James and John Dewey. Pragmatism is concerned with actions and beliefs, where our actions are determined by our beliefs and in turn, our beliefs will be influenced by the outcome of our actions (Morgan, 2014b). Thus, action and beliefs can never be separated from each other. Dewey proposed that this happened automatically to produce habits but could happen consciously to create what Dewey termed inquiry and is at the heart of research underpinned by pragmatism (Dewey, 2008). Dewey proposed five steps of inquiry that can be used for research. Recognising the research problem; reflecting on ways to deal with the research problem; developing possible actions to address the research; evaluating these actions to see what the consequences might be and finally carrying out the research. The first four steps can be repeated until the research is carried out (Morgan, 2014a; Morgan, 2014b).

Pragmatism bridges the gap between the two extremes of philosophical thought, post-positivism and constructivism, taking on board both views (Creswell and Creswell, 2018). Pragmatists believe there is an objective reality, but this is experienced individually (Morgan, 2014a). Although pragmatism is most commonly associated with mixed methods research (Glogowska, 2015; Teddlie and Tashakkori, 2009), for a pragmatist it does not matter whether the research design is qualitative, quantitative, or mixed methods. The most important thing to consider is what the most appropriate method is for the research question (Morgan, 2014b). Pragmatism 'assumes an independence of methods' (Teddlie and Tashakkori, 2009).

Dewey (2018) suggested that all scientific inquiry should grow out of social conditions. Pragmatism is thus interested in solving social problems that lead to action for improving practice (Kaushik and Walsh, 2019). As a speech and language therapist trained in the social model, pragmatism aligns with my values. Both evidence-based research and a person-centred approach are fundamental in achieving successful therapy. The person-in-environment is an important concept in pragmatism. Speech and language therapists cannot understand or help their clients fully without knowledge of their environmental context. The way a speech and language therapist carries out therapy is very similar to Dewey's inquiry model where recognising a problem, reflection, evaluation and considered action is all important. Thus, this research, although it is quantitative in design, is interpreted and aligned with a pragmatist perspective.

1.6 Rationale for the study

The rationale for undertaking the study is there is little information on the association between wellbeing and voice symptoms. Additionally, there is still lack of clarity around which factors are associated with voice problems. This is largely due to methodological limitations of the studies that the literature review will address. In studies with teachers only Akinbode et al. (2014) addresses the possible clustering of teachers in the same school and thus school effects, in their statistical models, so their results may not be true associations. Clustering refers to the possibility that teachers working at the same school may be similar in certain respects and needs to be accounted for during the analysis. This study will therefore use a multilevel model to address this and will be the first study in the voice risk factor research with teachers to use this statistical approach.

Additionally, there has been little risk factor research undertaken with teachers in the UK. The results of studies in other countries may not be generalisable due to factors such as different teaching methods, cultural and social differences including possible differences in reporting voice problems and the perception of voice problems. There have been two studies identified that were carried out in the UK. These include a study in Northern Ireland (McAleavy et al., 2008) and the Northwest of England (Gadepalli et al., 2019). However, neither of these account for school effects in their statistical models or investigates wellbeing, the focus of this study. This study decided to have a nationwide focus within England so that results could be as generalisable as possible to teachers in the whole of the country. The rest of the countries within the UK were not included due to differences in the organisation of schools, curriculum and assessment of children between England, Wales, Scotland and Northern Ireland. In particular, Scotland has a different qualification framework than the rest of the countries in the UK (Education Scotland, 2018). Differences in the education systems in different countries within the UK mean that teachers from each country may not be comparable.

Wellbeing is the primary focus of the study, as it has not been investigated in the voice literature before. Although psychological problems have been included in other studies, as wellbeing is a separate construct, albeit linked, it is hoped that the results will provide a useful addition to the literature.

1.7 Aim of the Study

- To investigate the link between wellbeing and voice symptoms in primary and secondary school teachers in England.

1.8 Objectives of the Study

- To determine the risk factors that are associated with voice symptoms in primary and secondary school teachers in England.
- Address the limitations of existing research.
- To add to the current knowledge and evidence base for teachers and health care professionals in order to help inform treatment and prevention strategies.

1.9 Overview of the Thesis

The next section of the thesis, chapter 2, presents the findings of the literature review that begins with outlining the search strategy for the papers that were included. It then appraises these studies in order to provide further rationale for the study. Chapter 3 presents the methodology, including recruitment of participants, questionnaire design, pretesting of the questionnaire and details of how the statistical analysis was conducted. This section also provides a discussion of the ethical considerations of the study. Chapter 4 provides the findings of the analysis, comparing results from a univariate model, a multiple linear regression and a multilevel model. The thesis concludes with chapter 5, which provides a discussion of the study including its limitations, suggestions for further research and applications for clinical practice and in schools.

Chapter 2. Literature Review

2.0 Introduction

This chapter investigates the literature regarding risk factors for voice problems in teachers working in primary and secondary schools in England. The literature review was undertaken in a systematic way, although it was not possible to carry out a full systematic review due to time, resources and the heterogeneity of the questions asked. Although the primary aim of the research is to investigate the association of wellbeing and voice problems, the research also investigates other factors. The chapter will outline the search strategy used to select papers and discuss the inclusion and exclusion criteria before reporting the results of the selected papers and critically appraising them. The chapter will also include a background section that will consider the evidence base for current risk factors and how applicable the studies are to the teaching population.

2.1 Background

An important part of the research into voice problems has been the study of their risk factors. This research has helped to inform preventative strategies and treatment approaches for voice problems. The evidence base for these risk factors comes from a variety of sources. They include cell biology studies using human cell cultures, bench studies using excised animal larynges, experimental studies with human participants and observational studies. This section briefly discusses each type of study in terms of their findings and their relevance to teachers with voice problems.

Cell culture studies have primarily investigated the role reflux has played on the larynx and looked at the biological properties of cells from patients with laryngopharyngeal reflux (LPR). These studies have found that reflux is harmful to laryngeal mucosa (Johnston et al., 2007; Ylitalo et al., 2004) and for the laryngeal cells of patients with LPR to have compromised laryngeal defence mechanisms (Gill et al., 2005; Johnston et al., 2003; Johnston et al., 2004; Samuels et al., 2008). Theoretically, this evidence suggests that reflux is likely to create voice problems. However, isolated cells taken from the human body can behave differently from when they are in an organism (Zellmer et al., 2010), which means findings from these studies can be difficult to extrapolate to real life situations.

Studies using excised larynges from dogs, sheep and pigs have investigated the properties of the vocal folds after dehydration and rehydration. Experiments have suggested that dehydration leads to increased stiffness and viscosity of the vocal folds (Chan and Tayama, 2002; Hemler et al., 2001; Miri et al., 2012) and increased phonatory threshold pressure (PTP) (Finkelhor et al., 1988; Jiang et al., 2000) and that these effects are reversed when the vocal folds are rehydrated. PTP refers to the minimum pressure below the vocal folds required to initiate phonation (Titze, 1988). Theoretically,

an increase in viscosity of the vocal folds and PTP should make voice problems more likely and therefore increasing hydration should reverse voice problems. However, the results from animal studies may not be generalisable to human populations. The larynges are researched independently from the rest of the organism and therefore, it is only possible to measure surface level hydration, known as superficial hydration of the larynx, rather than considering the hydration of the whole body (systemic hydration). The experiments also artificially dehydrate and rehydrate the larynges. For example, Chan and Tayama (2002) immersed canine larynges in hypertonic and hypotonic solutions and Hemler et al. (2001) placed sheep larynges in 0% humidity. However, it is unlikely, under normal circumstances, and considering homeostasis, that the human body would be subjected to the levels of dehydration performed in these experiments. Therefore, the results of these experiments can only be considered theoretical. In order to make these findings applicable to real life, research with human participants needs to be considered.

Experimental studies investigating dehydration have been carried out with human participants, using PTP and perceived phonatory effort (PPE) as outcome measures. Some of the research has investigated superficial hydration/dehydration (Levendoski et al., 2014; Tanner et al., 2007) and others have tested the effects of systemic hydration (Verdolini-Marston et al., 1990; Verdolini et al., 2002; Verdolini et al., 1994) on vocal functioning. The experiments found that PTP increased following dehydration, therefore concluding that voice problems may increase. However, in these experiments, PTP levels were only found to be significant at high pitches and therefore, results may not be applicable to conversational levels of pitch. Indeed, PPE levels, which were measured during normal conversation, did not always correspond to an increase in PTP. For example, Verdolini et al. (2002) found PPE did not increase following dehydration with a diuretic, and Tanner et al. (2007) found PPE decreased after dehydration. As with all experiments, ecological validity is an issue. The experiments are taking part in laboratories and the conditions they are presented with, are often not applicable to normal life situations. Participants are artificially dehydrated either by reducing the humidity in the room or by providing diuretics. For example, Verdolini et al. (1994) asked participants to take two teaspoons of a decongestant and abstain from fluids for the dehydration condition, and take a mucolytic drug, a drug which helps with the clearance of mucus, and drink as much water as they could tolerate for the rehydration condition. All this took part in an 'environmental chamber' at a temperature of 27 degrees Celsius. The results can therefore only be considered theoretical. In order to generalise results to real life, evidence from observational studies need to be evaluated.

Observational studies have been carried out with patients with certain health conditions and controls to compare self-assessed vocal functioning or prevalence of voice problems. Significantly

greater scores on the VHI have been found for patients with allergies and allergic rhinitis (Krouse et al., 2008; Millqvist et al., 2008; Özbal Koç et al., 2014; Turley et al., 2011), asthma (Dogan et al., 2007), third trimester of pregnancy (Saltürk et al., 2016), Sjögren's syndrome (Saltürk et al., 2017) and fibromyalgia (Gurbuzler et al., 2013) than controls. Additionally, studies have found a greater prevalence of voice problems among patients with rheumatoid arthritis (Speyer et al., 2008) and autoimmune disease (Liu et al., 2012) than controls. However, there are problems with the designs of some of these studies that mean the results may not be generalisable to the teaching population.

The selection of cases and controls in observational studies is often problematic and can lead to bias (Mann, 2003). Many of the studies cited above have possible biases due to the selection of their participants. For example, Gurbuzler et al. (2013) and Saltürk et al. (2016) selected controls without current voice problems, thereby causing a bias towards higher VHI results for the participants with health conditions. Krouse et al. (2008) did not report the age or the gender of participants and it may have been age or gender that accounted for the differences in VHI scores. Özbal Koç et al. (2014) did not report how they selected controls and therefore they may not have been sufficiently matched to the cases. As Mann (2003) states, controls should have similar risk factors and confounders to minimise bias. Although these studies provide evidence to suggest certain health factors may be associated with voice problems, they may not be generalisable to teachers. Teachers have a range of risk factors that are unique to their profession that need to be accounted for in a study. Therefore, in order to look at risk factors relevant to teachers, large studies which include teachers need to be evaluated. The rest of the literature review will critically analyse observational studies with teachers that investigate the association of risk factors with voice problems.

2.2 Aims and Objectives of the review

The literature review will help provide a rationale for the methodology of the study.

- The primary aim is: What is the association between voice symptoms and wellbeing in teachers?
- The secondary aim is: What other risk factors are associated with voice symptoms in teachers?

2.3 Methods

2.3.1 Criteria for inclusion and exclusion of studies

Table 3 shows the inclusion and exclusion criteria for studies

Table 3: Inclusion and Exclusion Criteria for Studies

	Inclusion Criteria	Exclusion Criteria
Participants	School teachers	Student teachers University lecturers
Schools	All levels and sectors of schools	
Voice problems	Current voice problems or had occurred in previous 12 months. These could be diagnosed by a doctor or self-reported	Self-reported voice problems over career of lifetime.
Country study took place in	All countries were considered	
Statistical analysis	Statistical model taking in account confounding variables such as multiple regression	Studies not taking account of multiple variables, e.g. simple regression.
Dates of publication	Year 2000 onwards	Before 2000
Literature	Peer reviewed	Grey literature.

Studies that were accepted included school teachers as the primary participants. They could be teaching in any level of school and all sectors, fee-paying or state education, mainstream or special education. Those with student teachers or university lecturers were excluded. This is because most student teachers and university lecturers will have different patterns of working, which means that some of the environmental risk factors they are subject to, will be different. For example, most student teachers will not have the same amount of contact time with pupils. University lecturers may not have as many contact hours with students as schoolteachers, and the age of the students will be older.

Studies were included that measured self-perceived voice problems that were current, or those that had occurred anywhere up to the last 12 months, or those that reported they had been diagnosed by a doctor. Studies that measured self-perceived voice problems over a longer period, for example over a participant's career or lifetime were not included. Problems with accurate recall are more likely over a longer period of time and therefore misclassification of a voice problem may occur (Webb and Bain, 2011). For example, if teachers are asked to report whether they have had voice problems over their career, recall bias is likely, with those who have recently joined the profession more likely to remember accurately than those who have been in the profession for many years.

Studies from all countries were considered. However, the search strategy identified only articles published in English. It is acknowledged there may be a potential language bias but due to the time and resource constraints needed for translation, it was not possible to search other literature. As this was a limited review only publications from 2000 onwards were considered. Only those from peer-reviewed journals were accepted. Due to time and resource constraints, the grey literature

was not searched and therefore it is acknowledged there may be publication bias. For example, McAuley et al. (2000) suggest that studies published in academic journals show larger effects when compared to grey literature.

Studies were only included that accounted for multiple variables. Those that only carried out simple linear regressions and thereby had not accounted for other variables, were excluded from the review. Models that do not account for multiple variables do not give an accurate picture of the association between the outcome and exposure variable. They only explain outcome in relation to one variable and ignore the possible impact of other variables. For example, a variable that is statistically significant in a univariate analysis may become statistically insignificant when other variables are taken into account, or indeed a variable that is unimportant in a univariate analysis may be statistically significant in a multiple regression model (Sun et al., 1996).

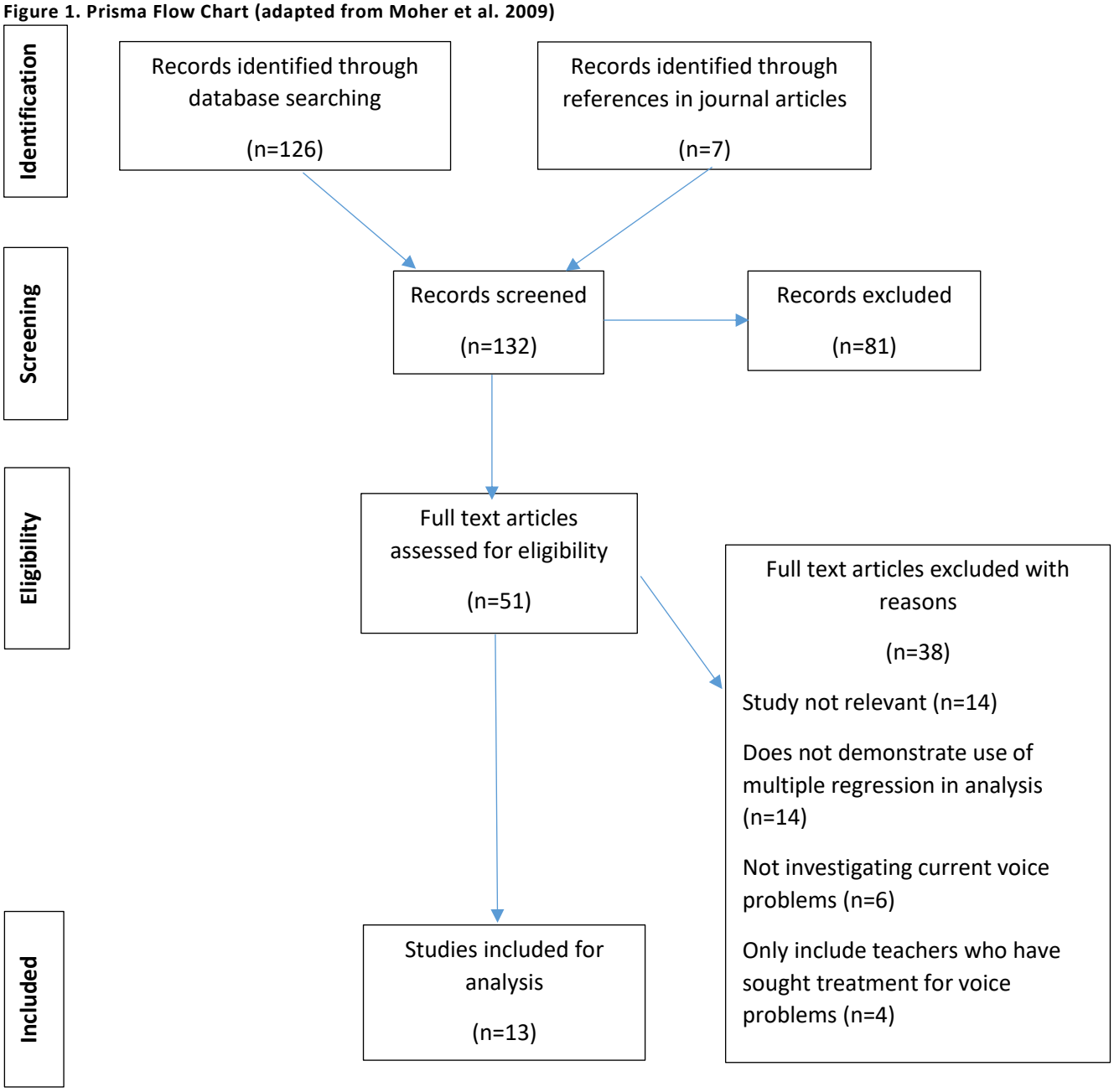
2.3.2 Literature Search

Electronic databases were used to search the literature. Those used were Medline, CINHAL, Psychinfo and Eric. The keywords used in the search were voice disorders, voice problems, dysphonia, teachers, risk factors, psychological wellbeing and mental health.

Three reviews were identified in the literature search: Martins et al. (2014b), Cutiva et al. (2013) and Byeon (2019). Cutiva et al. (2013) and Martins et al. (2014b) are reviews published over five years ago and therefore do not include the latest research. Additionally, the objectives for the review by Cutiva et al. (2013) are different as they are only interested in work-related factors such as working environment, employment conditions and voice use, not including health or lifestyle factors. Furthermore, all the reviews include papers that would not be accepted for this review. For example, they accept papers that look at voice problems over a lifetime, rather than just being interested in current voice problems, and select articles that do not take account for multiple variables and possible confounding variables in their statistical analyses. Both of these, as discussed earlier in the chapter, may suggest associations that are inaccurate. Therefore, a new review is needed.

A screening process of journal articles took place, applying the inclusion criteria. The outline of this is shown in figure 1. The titles and abstracts of papers identified through databases and through references in relevant journal articles (n=132) were screened for eligibility. As a result of this, 51 full text articles were assessed. Thirty eight papers were rejected with the following reasons: subject not relevant (n=14), did not demonstrate use of multiple regression in analysis (n=14), not investigating current voice problems (n=6), only including teachers who have sought treatment for voice problems (n=4). Following the screening process thirteen studies remained for inclusion in the review.

Figure 1. Prisma Flow Chart (adapted from Moher et al. 2009)



Data were extracted from articles that fulfilled the inclusion criteria. This included:

- Teaching level of the teachers
- Gender or sex
- Country the study took place in
- Sample size, and whether there was a sample size calculation
- How the participants were recruited when data collection took place

- How the study defined voice disorders
- Whether or not the study included mental health as a risk factor
- The dependent variables included in the study
- The results of the multiple regression analysis, and details about the statistical method.

2.4 Results

As a result of the search, thirteen studies that fitted the inclusion criteria were included in the review. The studies and their characteristics are outlined in table 4.

2.4.1 Study characteristics

This section will discuss the date of the study, the country it took place in, the teaching population the level of schools, the type of school, the sample size included in the analysis, how voice problems were measured and the response rate.

Table 4: Study Characteristics

Study (year and country)	Teaching population	School-level	Type of school	Sample size of teachers (included in analysis)	Measurement of voice problems	Response rate
Akinbode et al. (2014) Nigeria	Male and female. Head and assistant head teachers excluded.	Primary	Public and private	341	Presence of at least one of the following: hoarseness, repetitive throat clearing, tired voice, straining to speak.	77%
Assuncao et al. (2012) Brazil	Male and female	Does not report	Does not report	649	'yes' to the question: 'Has a doctor told you that you have dysphonia?'	Does not report
Charn and Mok (2012) Singapore.	Male and female. Full-time teachers.	Primary	Public schools	214	'Yes' to the question: 'Do you have a problem with your voice today which is preventing you from doing all that you want with it?'	Does not report
da Rocha et al. (2015) Brazil	Male and Female. PE teachers excluded.	Elementary	Municipal schools	575	VHI – score of 19 or over	90.8%
da Rocha et al. (2017). Brazil	Male and Female. PE teachers excluded.	Elementary	Municipal schools	469	VHI - score of 19 or over	81.56%
da Rocha and Souza (2013). Brazil	Male and female. PE teachers excluded.	Elementary	Municipal schools	575	VHI – continuous score used.	90.8%
de Alvear et al. (2011) Spain	Male and Female	Preschool and Elementary	Public schools	284	Vocal effort and a minimum of two vocal symptoms	28%
de Medeiros et al. (2008) Brazil	Female teachers only. PE and male teachers excluded.	Elementary	Municipal schools.	2103	'Have you felt too tired to speak in the last 2 weeks?' 'Have you perceived any loss of voice quality during the last 2 weeks?' Possible dysphonia – 'no' to one or 'sometimes' to both questions.	86%

					Probable dysphonia – 'daily' to one or both questions.	
Gadepalli et al. (2019) North West England	Male and female teachers.	Primary and secondary	Mainstream schools	454 teachers 304 non teachers	VHI-10 Used continuous score	40%
Lee et al. (2010) Hong Kong	Male and female	Primary.	Mainstream schools. Exclude special education	498	'yes' to the question: 'For the purposes of this study, we consider a voice disorder to be any time your voice does not work, perform or sound as you feel it usually does, so that it interferes with communication. Have you had voice disorder like this in the last 12 mo?'	69.7%
Moy et al. (2015) Malaysia	Male and female. Permanent employment. Free of Mental health problems	Secondary	Public schools	6039	VHI 10 – score of greater than 11	88.1%
Rossi-Barbosa et al. (2016) Brazil	Female teachers. Male and PE teachers excluded.	Elementary	Municipal	226	'Have you noticed changes in your voice quality?' Acute – 3 weeks or less Chronic – more than 3 weeks.	Do not report
Sampaio et al. (2012). Brazil	Male and female, temporary and permanent staff	Elementary (and preschool?).	Municipal schools	4496	VHI 10 Minimal handicap: below 11 Moderate/severe handicap: 11 and over	95.7%

Thirteen studies were included, published between 2008 and 2019. Most of the studies were carried out in Brazil, although three of these (da Rocha et al., 2017; da Rocha et al., 2015; da Rocha and Souza, 2013) used the same population of teachers. The da Rocha et al. (2017) study followed up the participants included in the da Rocha et al. (2015) and da Rocha and Souza (2013) studies to provide longitudinal data. Three of the studies took place in South East Asia, one in Nigeria and one in Spain. There was only one study identified, carried out in the UK that fitted the inclusion criteria. This indicates a paucity of robust research that has been carried out in UK and indeed Europe. Studies from other countries may not be generalisable to teachers in the UK due to different teaching systems and cultural differences. Therefore, more well-designed UK based studies would be beneficial.

2.4.1.1 Teaching population

Most of the studies report that they include male and female teachers although Rossi Barbosa et al. (2016) and de Medeiros et al. (2008) exclude male teachers. Charn and Mok (2012) recruited full time teachers, Moy et al. (2015) included those in permanent employment while Sampaio et al. (2012) included both temporary and permanent teachers. Da Rocha and Souza (2013), da Rocha et al. (2015), da Rocha et al. (2017), de Medeiros et al. (2008) and Rossi-Barbosa et al. (2016) all

excluded PE teachers. The remainder of the studies do not report the eligibility criteria of their teaching population. The rationale for excluding male teachers in the de Medeiros et al. (2008) study were differences in prevalence rates of voice problems between men and women and in the Rossi-Barbosa et al. (2016) study, differences in the structure of the larynx. Both studies also gave a rationale of their being few male teachers in the school system. However, the evidence for prevalence rates being higher amongst male teachers is inconclusive. Not including males in the study misses out an important factor (gender/sex) that could add knowledge to the evidence base. The rationale for excluding PE teachers was differences in vocal demands and teaching activities to other subjects. Excluding PE teachers may miss out important information on risk factors potentially related to subject taught which could enhance understanding. It would be useful to include PE teachers in order to see if they are more likely to have voice problems, as research suggests that they have greater vocal load than teachers of other subjects (Nusseck et al., 2018).

2.4.1.2 Type of school

The majority of the studies recruit teachers from state funded schools although Akinbode et al. (2014) also include teachers from private schools. Gadepalli et al. (2019) and Lee et al. (2010) exclude special schools. The remainder of studies do not report whether special schools were included or excluded and thus it is not possible to know whether results can be generalised to teachers from these schools. Thorough reporting of the methodology of studies is essential. As Vandenbroucke et al. (2014, 1500) state in their guideline Strengthening the Reporting of Observational Studies in Epidemiology (STROBE), "poor reporting hampers the assessment of the strengths and weaknesses of a study and the generalisability of its results".

2.4.1.3 School level

Three studies involved teachers at primary schools (Akinbode et al., 2014; Charn and Mok, 2012; Lee et al., 2010), one study included participants from secondary schools (Moy et al., 2015) and one study included both primary and secondary schools (Gadepalli et al., 2019). Four of the Brazilian studies included teachers from elementary schools, which take pupils from 6 to 14 years of age (da Rocha et al., 2017; da Rocha et al., 2015; da Rocha and Souza, 2013; de Medeiros et al., 2008) and de Alvear et al. (2011) included teachers from preschool and elementary schools. In the Sampaio et al. (2012) study it is not possible to determine whether the teachers taught at elementary schools and preschools or elementary schools only, as the abstract and methodologies report different information. Assuncao et al. (2012) did not report school-level, so it is not possible to generalise the findings from this study to a particular level of school.

2.4.1.4 Measuring voice problems

Many of the studies use methods of measuring a voice problem that they have devised themselves. A common approach to assess whether someone has a voice problem was to use a question where a binary yes or no response was required. Akinbode et al. (2014) and de Alvear et al. (2011) present their participants with a list of symptoms. Akinbode et al. (2014) classify a voice problem as a participant reporting the presence of at least one voice symptom and de Alvear et al. (2011) as vocal effort and a minimum of two vocal symptoms. Charn and Mok (2012) and Lee et al. (2010) use a question based on one first used in a study by Roy et al. (2004, 283) measuring voice problems in teachers in the USA "For the purposes of this study, we consider a voice disorder to be any time your voice does not work, perform, or sound as you feel it normally does, so that it interferes with communication. Have you ever had a voice disorder like this?" although Lee et al. (2010) asks teachers about their voice in the last 12 months and Charn and Mok (2012) about current voice problems. De Medeiros et al. (2008) classify 'probable' dysphonia as participants answering daily to one or both questions regarding vocal tiredness and loss of voice quality and Rossi-Barbosa et al. (2016, 755.e26) determines a voice problem as participants reporting yes to the question "Have you noticed changes in your voice quality?" All but one of the remainder of the studies use validated measures. Da Rocha and Souza et al. (2013), da Rocha et al. (2015) and da Rocha et al. (2017) use the VHI and Gadepalli et al. (2019), Moy et al. (2015) and Sampaio et al. (2012) use the VHI 10.

The study by Assuncao et al. (2012), asked participants to report whether they had received a diagnosis of a voice disorder from a doctor or ENT. Although this is likely to result in accurate classification of voice problems, it would miss participants in the study who had a voice problem but had not been to see their doctor. This is likely to result in a reduced estimate of prevalence but also result in bias. Those who visit the doctor with a voice problem may be different from those who do not visit the doctor in ways that may affect the results of the association. For example, research suggests that those who are most likely to consult their doctors are more likely to have conditions increasing in severity or think stress is associated with their illness (Campbell and Roland, 1996).

The studies that don't use validated measures and those that have undergone psychometric testing are more likely to incorrectly measure voice problems. Using an inappropriate tool to measure voice problems will affect the validity of these studies and thus the strength of associations in these studies may be incorrect. Furthermore, using a single sentence question requiring a binary answer or asking participants to choose from a list of symptoms does not capture the breadth and complexity of a voice problem and could lead to misclassification of the outcome variable. Voice disorders are multifactorial and can involve a range of symptoms and therefore any self-report tool needs to reflect this complexity.

The WHO has developed the International Classification of Functioning Disability and Health (ICF) (World Health Organisation, 2001) which focuses on the impact that a health condition has on someone's day-to-day life. The ICF's subjective constructs of activity and participation have highlighted the importance of considering not only a person's symptoms but also how they can function with their symptoms. Therefore, any outcome measure should ideally incorporate questions related to functioning, rather than just physical symptoms. There are validated self-administered tools used for the voice that take into account, not only symptoms but the daily functioning of a person including the emotional impact of a voice problem. These are tools such as the VHI (Jacobson et al., 1997), VoiSS (Deary et al., 2003) and VPQ (Carding et al., 1999). They were originally designed for use in clinical situations as part of a clinical voice assessment. However, they have more recently been used in research to classify voice problems. In order to achieve this, researchers have used cut off scores, so that a score above a certain value indicates a voice problem or voice disorder.

Teacher risk factor studies that use cut off scores include da Rocha et al. (2015) and da Rocha et al. (2017) with the Brazilian version of the Voice Handicap Index (VHI), classifying a voice problem as a score of 19 or above. Moy et al. (2015) classifies a voice problem as a score greater than 11 on the VHI-10 and Sampaio et al. (2012) regards a score of 11 and over as a moderate/severe handicap and a score of below 11 as a minimal handicap. Although it is useful in creating a binary outcome, creating cut off scores is problematic. Firstly, there is no official gold standard for diagnosing voice disorders which a scale can be measured against, for specificity and sensitivity. The sensitivity of an instrument is the probability that someone who has the disease will be diagnosed. Specificity refers to the probability that someone who does not have the disease will be falsely diagnosed (Webb and Bain, 2011). Secondly, dichotomising an instrument designed to be used continuously loses information, so the power to detect an association between an independent variable and the dependent variable is reduced (Altman and Royston, 2006).

If prevalence is the primary aim of the study, then using a validated instrument with a cut off score, may be the most valid and reliable method to determine a voice problem. It would be essential though that the study uses cut off scores that are suitable for its population. For example da Rocha et al. (2015) and da Rocha et al. (2017) use cut off scores for the Brazilian Portuguese version of the VHI from a study by Behlau et al. (2015) using the Brazilian Portuguese version. Therefore, the cut off score used in these studies is more likely to classify voice disorders correctly. However, Sampaio et al. (2012) do not report where they got their cut off score from, and Moy et al. (2015) uses a cut off score based on normative values from a study carried out with a US population (Arffa et al.,

2012) which may not be valid for a Malaysian teaching population. Therefore, the studies by Moy et al. (2015) and Sampaio et al. (2012) may not be reliably classifying voice problems.

Using cut off scores may be problematic in other ways. Someone just below the cut off score may consider themselves to have a voice problem. Therefore, these studies may not be reliably classifying voice problems in all situations. Perhaps the only valid way to look at associations between voice problems and risk factors is to use validated instruments as a continuous measure as do Gadepalli et al. (2019) in their multivariate regression and da Rocha and Souza (2013). Additionally, voice clinics look at a client's total score rather than using cut off scores, and thus results from using the instruments as a continuous measure may be more applicable to clinical situations.

2.4.1.5 Sample size

The sample sizes of the studies range from 214 (Charn and Mok, 2012) to 6039 (Moy et al., 2015). These numbers reflect the participants that took part in each study, rather than the number of participants that the studies calculated or considered were needed. There are limitations with the way some of the studies determine their required sample sizes, with four of the studies not reporting their considerations (Akinbode et al., 2014; Charn and Mok, 2012; de Alvear et al., 2011; Moy et al., 2015). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) (Vandenbroucke et al., 2014, 511) gives recommendations on how observational studies should report on how they have determined their sample size. They "encourage investigators to report pertinent formal sample size calculations if they were done. In other situations, they should indicate the considerations that determined the study size". Sample sizes need to be large enough in order to reduce the likelihood of type II errors occurring (Jackson and Smith, 2013) and also in the case of risk factor studies, to detect an association between an outcome variable and independent variables.

Over half of the studies carry out a sample size calculation in order to determine the sample size of the study. (Assuncao et al., 2012; da Rocha et al., 2017; da Rocha et al., 2015; da Rocha and Souza, 2013; De Medieros et al., 2008, Lee et al., 2010; Rossi-Barbosa et al., 2016). However, as these studies are observational rather than RCTs and are hypothesis generating rather than confirming, they typically won't have an effect size from which to draw their sample. Some of the studies calculate their sample sizes based on prevalence of voice problems (Assuncao et al., 2012; De Medieros et al., 2008; Lee et al., 2010; Rossi-Barbosa et al., 2016), using predicted prevalence, confidence levels and precision in their calculations (Lwanga et al., 1991). However, as all of these studies are investigating associations of voice problems and risk factors using multiple regression analyses, which require different considerations for sample size, the prevalence calculations may not

be sufficient, and therefore type II errors may still occur. Furthermore, de Medeiros et al. (2008) do not report what elements they used in the sample size calculation, so it is not known if an appropriate mathematical formula was used. Additionally, they did not report the result of their calculation, so it is not possible to determine whether the number of participants they recruited was large enough to avoid Type II errors.

Two of the studies carry out a census-based approach, asking all teachers in the study population to take part (Gadepalli et al., 2019; Sampaio et al., 2012). However, neither of these studies provide a rationale for this approach. The Sampaio et al. (2012) study in particular had a very high response rate of 95.7% with a sample size of 4496. Indeed, this may mean that the sample size is too large which may have ethical implications, such as exposing a large number of participants to unnecessary burden (Jackson and Smith 2003).

2.4.1.6 Response Rate

Response rates for the studies ranges from 28% (de Alvear et al., 2011) to 95.7% (Sampaio et al., 2012) with three of the studies not reporting their response rates (Assuncao et al., 2012, Charn and Mok 2012, Rossi-Barbosa et al 2016). Two of the studies had low response rates (de Alvear et al., 2011, Gadepalli et al., 2019). A high response rate is generally considered positive as it often lowers the chance of response bias (Groves, 2009). Response bias indicates there is a difference between those who respond to the survey and those who do not and can reduce a study's generalisability (Aday and Cornelius, 2006). However, a high response rate does not always indicate that the study's population is representative of the population it is to be generalised to. Indeed the representativeness of a study population is often considered more important than a high response rate (Cook et al., 2000). In order to demonstrate that a study has a representative sample a study needs to compare the proportions of teaching characteristics reported in nationwide teacher censuses and compare them with those in the study. Only Charn and Mok (2012) and de Alvear et al. (2011) report this. Therefore, even with the studies with high response rates it is not possible to determine how representative their results are.

2.4.2 Study Outcomes

2.4.2.1 Outcomes for associations between psychological factors and voice problems.

Seven of the studies measure aspects of mental health. These results are displayed in table 5.

Table 5: Mental Health

Study	Mental Health problem	How MH is measured	Result
Assuncao et al. (2012)	Common Mental Disorder Job related stress	Brazilian version of Self-Reporting Questionnaire Short version of Job-related Stress Scale (Portuguese version)	No association No association
da Rocha et al. (2017)	Common Mental Disorder	Brazilian version of Self-Reporting Questionnaire	Relative Risk (RR) 2.09 (95% CI 1.65, 2.64)
da Rocha and Souza (2013)	Common Mental Disorder	Brazilian version of Self-Reporting Questionnaire	ß0.279 (0.210, 0.347)
da Rocha et al. (2015)	Current depressive episode	Mini International Neuropsychiatric Interview	Prevalence Ratio (PR) 1.66 (1.31, 2.10)
de Medeiros et al. (2008)	Mental disorder	Brazilian Portuguese version of General Health Questionnaire	Odds Ratio (OR) 2.40 (1.90, 3.03) possible and probable dysphonia OR 3.20 (2.18, 4.70) probable dysphonia
Gadepalli et al. (2019)	Job related stress	Single question	No association
Moy et al. (2015)	Anxiety, depression and Stress	Malay version of Depression, Anxiety and Stress Scale 21	Anxiety PR 1.04 (1.02, 1.06) Depression PR 1.02 (0.99, 1,04) no association Stress PR 1.00 (0.96, 1.06) no association.

Five of the studies show a statistically significant relationship between aspects of mental health and voice problems. da Rocha et al. (2017) and da Rocha and Souza (2013) found a statistically significant association with common mental disorder and voice problems, da Rocha et al. (2015) found a significant association with current depressive episode and voice problems, Moy et al. (2015) found a significant association between anxiety and voice problems and de Medeiros et al. (2008) found a statistically significant association between mental disorder and voice disorders.

A strength of many of these studies is that they use validated tools to measure mental health (Assuncao et al., 2012,; da Rocha and Souza, 2013; da Rocha et al., 2015; da Rocha et al., 2017; de Medieros et al., 2008; Moy et al., 2015), indicating that the classification of mental health problems is likely to be valid and reliable. A small width of confidence intervals in the studies, indicates precision in the results. However, there are some particular limitations in some of the studies which may indicate that their results are biased. For example, as part of their exclusion criteria Moy et al. (2015) report that they excluded participants with mental health problems, although there is no explanation of what they considered a mental health problem to be. Therefore, their participants are likely to be biased towards those with mild mental health problems and thus results may not show a true association between mental health and voice problems. Assuncao et al. (2012) found no relationship between common mental disorder or job-related stress and voice problems.

However, in their study they only classify voice problems when participants have reported having a diagnosis from their doctor. There may be a difference in the mental health between those who seek treatment for voice problems, and those who have voice problems but do not seek treatment. It is possible that if the authors had included participants with voice problems regardless of whether they had sought treatment the results would have been positive. Gadepalli et al. (2019) do not use a validated tool to measure job-related stress and therefore misclassification may have occurred.

2.4.2.2 Outcomes for other risk factors and voice problems.

Table 6 shows the associations of other risk factors and voice problems in the selected studies after multiple regression. This section will look at the similarity of results across studies to see if conclusions can be drawn about any of the risk factors and their associations with voice disorders. The results between different types of measurement are difficult to compare. Odds ratios and prevalence rates are measuring different things. The estimates of odds ratios and prevalence rates are only comparable when the prevalence of a condition is rare, authors commonly quote 10% or under (Barros and Hirakata, 2003; Behrens et al., 2004; McNutt et al., 2003).

Table 6: Associations of Risk Factors and Voice Problems after Multiple Regression

(Statistically significant results in bold)

Variable	Study	Result
Gender/Sex		
Male gender	Akinbode et al. (2014)	OR 0.62 (0.27, 1.46)
Female sex	Assuncao et al. (2012)	**PR 2.33 (1.41, 3.85)**
	Sampaio et al. (2012)	**PR 1.72 (1.20, 2.48)**
Female gender	de Alvear et al. (2011)	**OR 3.56 (1.49, 8.49)**
	Charn and Mok (2012)	No association
Age		
35-48 (ref below 35)	Akinbode et al. (2014)	OR 1.44 (0.55, 3.78)
Age 49-60		OR 1.62 (0.45, 5.82)
</=29 years (ref >/=50)	Moy et al. (2015)	OR 1.09 (0.47, 2.52)
30-39		OR 1.13 (0.94, 1.35)
40-49		**OR 1.20 (1.09, 1.32)**
decades	de Alvear et al. (2011)	OR 1.19 (0.81, 1.75)
Vocal Effort		
Talks a lot or excessively	Rossi-Barbosa et al. (2016)	**OR 3.72 (1.48, 9.40) acute voice problem**
		OR 3.36 (1.16, 9.78) chronic voice problem
Vocal behaviour profile	Da Rocha et al. (2015)	
'Serious risk of abuse'		PR 2.58 (0.99, 6.73)
'Champion of abuse'		**PR 5.33 (2.07, 13.68)**
Raised voice while teaching	Akinbode et al. (2014)	**OR 10.1 (5.07, 20.2)**
Work in another activity with intense use of voice	De Medeiros et al. (2008)	OR 1.13 (0.83, 1.54) possible and probable dysphonia
		OR 1.71 (1.08, 2.71) probable dysphonia
Professional vocal effort (LVEI)	Sampaio et al. (2012)	**PR 1.47 (1.19, 1.82)**
Speaking against background noise – frequently	Lee et al. (2010)	**OR 1.8 (1.1, 2.8)**
Attributing vocal problems to prolonged vocal use	de Alvear et al. (2011)	**OR 5.62 (2.42, 13.04)**
Noise		
Noise generated within the classroom (high to unbearable)	de Medeiros et al. (2008)	OR 1.30 (0.95, 1.51) possible and probable dysphonia
		OR 2.55 (1.72, 3.76) probable dysphonia
Perception of noise outside the school (disturbing to unbearable)	Rossi-Barbosa et al. (2016)	OR 1.23 (0.36, 4.16) acute voice problems
		OR 4.23 (1.29, 13.85) chronic voice problems
Excessive noise	Sampaio et al. (2012)	**PR 1.47 (1.28, 1.70)**

Health Conditions		
Upper airway problems (past 15d)	De Medeiros et al. (2008)	**OR 3.16 (2.47, 4.05) Possible and Probable Dysphonia** **OR 5.95 (4.06, 8.77) Probable dysphonia**
Upper respiratory tract infection often	Akinbode et al. (2014)	**OR 3.60 (1.39, 9.33)**
Laryngitis 1-5 times	Lee et al. (2010)	**OR 2.0 (1.1, 3.8)**
Laryngitis >5 times		**OR 4.2 (2.1, 8.5)**
Heartburn	Sampaio et al. (2012)	**PR 1.74 (1.50, 2.02)**
Reflux	Charn and Mok (2012)	**OR 6.09 (2.54, 14.52)**
Rhinitis	Sampaio et al. (2012)	**PR 1.35 (1.17, 1.55)**
Asthma	Lee et al. (2010)	**OR 3.3 (1.4, 8.3)**
Cough	Gadepalli et al. (2019)	**Incidence rate 1.455**
Diagnosis of gastritis	Assuncao et al. (2012)	**PR 1.59 (1.28, 1.98)**
Absenteeism due to illness or sick leave	Assuncao et al. (2012)	**PR 1.39 (1.06, 1.81)**
Perceived state of health below excellent levels	de Alvear et al. (2011)	**OR 0.58 (0.34, 0.97)**
Needing a long rest to relieve voice symptoms	de Alvear et al. (2011)	**OR 4.97 (2.88, 8.56)**
Lifestyle Factors		
Regular physical activity (1-2 times a week)	de Medeiros et al. (2012)	**OR 1.36 (1.01, 1.84) possible and probable dysphonia**
Regular physical activity (none)		OR 1.26 (0.75, 2.12) probable dysphonia Or 1.30 (0.98, 1.73) possible and probable dysphonia **OR 1.93 (1.21, 3.10) probable dysphonia**
Water intake per day (<4 glasses)	Rossi-Barbosa et al. (2016)	**OR 2.26 (1.14, 4.49) acute voice problem** OR 1.43 (0.65, 3.15) chronic voice problem
Consumption of units of alcohol at a time (>1U)	Rossi-Barbosa et al. (2016)	**OR 2.19 (1.07, 4.49) acute voice problem** **OR 3.19 (1.41, 7.25) chronic voice problem**
Alcohol intake occasionally	Akinbode et al. (2014)	OR 0.75 (0.33, 1.70)
Alcohol intake at least weekly		OR 5.01 (0.49, 50.8)
Alcohol consumption frequently	Lee et al. (2010)	**OR 0.4 (0.2, 0.8)**
Caffeinated drink intake regularly	Akinbode et al. (2014)	**OR 3.17 (1.51, 6.62)**
Habit of singing occasionally	Akinbode et al. (2014)	OR 0.48 (0.23, 1.52)
Habit of singing daily		**OR 0.48 (0.23, 0.99)**
Smoked > 5 in the last 5 y	Gadepalli et al. (2019)	Incidence Rate 0.977
Frequent use of mints or balm sprays	de Alvear et al. (2011)	OR 2.37 (0.92, 6.09)
Environmental factors in the classroom		
Ventilation in the classroom (reasonable)	de Medeiros et al. (2008)	**OR 1.65 (1.28, 2.14) possible and probable dysphonia**
Ventilation in the classroom (poor)		OR 1.37 (0.88, 2.14) probable dysphonia OR 1.30 (0.96, 1.77) possible and probable dysphonia **OR 2.00 (1.24, 3.22) probable dysphonia**
Perception of technical resources and equipment in workplace	Assuncao et al. (2012)	
Fair		PR 1.21 (0.93, 1.59)
Poor		**PR 1.56 (1.14, 2.15)**
Number of students per class (26 or more)	da Rocha et al. (2017)	RR1.09 (0.85, 1.39)
	da Rocha et al. (2015)	PR 1.25 (0.97, 1.61)
Class size 25-35	Akinbode et al. (2014)	OR 0.52 (0.22, 1.27)
Class size 36-94		OR 0.63 (0.18, 2.22)
Years teaching (>15 years)	Rossi-Barbosa et al. (2016)	OR 1.61 (0.7, 3.30) acute voice problem **OR 3.10 (1.28, 7.48) chronic voice problem**
Years teaching 8-20 years	Akinbode et al. (2014)	OR 0.83 (0.31, 2.23)
Years teaching 21-35		OR 0.63 (0.18, 2.22)
Amplifier use – yes	Moy et al. (2015)	PR 1.84 (0.89, 3.81)
	Lee et al. (2010)	OR 1.5 (1.0, 2.4)
Using microphone during teaching frequently	Sampaio et al. (2012)	**PR 1.69(1.32, 2.17)**
No microphone in the classroom	Charn and Mok (2012)	
Pressure from school management	Sampaio et al. (2012)	**OR 5.61 (1.79, 17.61)**
Attributing vocal problems to indiscipline	de Alvear et al. (2011)	**PR 1.22 (1.04, 1.44)**
Attributing vocal problems to teaching PE or a foreign language	de Alvear et al. (2011)	**OR 2.48 (1.14, 5.37)** OR 2.45 (0.92, 6.54)

The variables most commonly associated with voice problems across the studies are those related to vocal effort. Six of the seven studies that included an aspect of vocal effort in their multiple regression studies found statistically significant associations with voice problems. These included:

- talking a lot or excessively (Rossi-Barbosa et al., 2016),
- 'champions of abuse' (participants most likely to use their voice in a damaging way) (da Rocha et al., 2015) measured using the Vocal Behavior Profile (Behlau et al., 2018),
- raised voice while teaching (Akinbode et al., 2014),
- speaking against background noise (Lee et al. 2010),
- working in another activity with intense use of voice (de Medeiros et al., 2008),
- higher lifetime vocal effort (number of years working as a teacher multiplied by the mean number of hours worked per week) (Sampaio et al., 2012)
- attributing vocal problems to prolonged vocal use (de Alvear et al., 2011).

A variable closely linked to vocal effort, is noise perceived within and without the classroom. According to the Lombard Effect, if teachers perceive the noise around them to be high, they will involuntary increase the effort required to speak (Lane and Tranel, 1971). Noise generated in the classroom (high to unbearable) (de Medeiros et al., 2008), perception of noise outside the school (disturbing to unbearable) (Rossi-Barbosa et al., 2016) and excessive noise (Sampaio et al., 2012) were all found to be significantly associated with voice problems. Therefore, there seems to be compelling evidence that voice problems are linked to vocal effort. However, some of the confidence intervals are wide for these associations indicating low precision in the estimates. These include: perception of noise outside the school (disturbing to unbearable) (95% CI 1.29, 13.85 chronic voice problems) and talks a lot or excessively (95% CI 1.48, 9.40 acute voice problem, 1.16, 9.78 chronic voice problem) (Rossi-Barbosa et al., 2016), 'Champion of Abuse'(95% CI 2.07, 13.68) (da Rocha et al., 2015), and raised voice while teaching (95%CI 5.07, 20.2) (Akinbode et al., 2014).

However, a variable, which could be considered to be linked to vocal effort, years working as a teacher, was only statistically significantly associated with voice problems in one study (Rossi Barbosi et al., 2016). It may be that the teachers who have worked for many years may have adapted their voice technique so that they use less effort to speak.

Another variable closely linked to vocal effort is microphone or amplifier use. It may be considered that using a microphone helps to prevent vocal problems. However, results between studies are not consistent. Sampaio et al. (2012) found that teachers who did not use a microphone in the classroom were at greater risk of having voice problems (PR 1.69 95% CI 1.32, 2.17), whereas Charn and Mok (2012) found that those who used a microphone had greater odds of having a voice problem (OR 5.61 95% CI 1.79, 17.61). It may be that teachers in the Charn and Mok (2012) study were more likely to use microphones if they had a voice problem. In the studies with Moy et al. (2015) and Lee et al. (2010), no association was found between amplifier or microphone and voice

problems. It is therefore, not possible to make any conclusions about microphone use and voice problems.

Gender/sex was found to be statistically significant in three out of five studies that included this variable in their multiple regression model. Assuncao et al. (2012), Sampaio et al. (2012) and de Alvear et al. (2011) found statistically significant associations between female sex/gender and voice problems. However, Akinbode et al. (2014) and Charn and Mok (2012) found no association. Charn and Mok (2012) however had a very small number of males in their sample (n=37), which may not have been large enough to show a difference. The remainder of the studies did not include the variable gender or sex in their statistical model either because they did not include males in their sample (de Medeiros et al., 2008; Rossi-Barbosa et al., 2016), or due to the methods for selecting variables for the final model. There are potential issues with the selection of variables for the final model in some of the studies, which will be discussed in section 2.4.3.2.

Many of the health conditions investigated across studies are different and show inconsistencies in results. Studies ranged from Moy et al. (2015) and the da Rocha studies not investigating any specific health conditions, to Lee et al. (2010) investigating the association of seven health conditions. The most common health conditions for studies to investigate were upper airway problems including laryngitis and respiratory infections, and reflux or heartburn. The most consistently significant health condition found to be related to voice problems were upper airway problems and infections (Akinbode et al., 2014; de Medeiros et al., 2008; Lee et al., 2010). de Medeiros et al. (2008) found those with upper airway problems in the past 15 days had significantly higher odds of having possible and probable dysphonia (OR 3.16 95%CI 2.47, 4.05) and probable dysphonia (OR 5.95 95%CI 4.06, 8.77). Akinbode et al. (2014) found those who often had an upper respiratory tract infection had significantly higher odds of having a voice disorder than those who seldom had upper respiratory tract infections (OR 3.60 95% 1.39, 9.33), although the CI was large indicating low precision. Lee et al. (2010) found those who had had laryngitis 1 to 5 times in the previous 12 months or more than 5 times had significantly higher odds of having a voice problem than those who had not had laryngitis (OR 2.0 95% CI 1.1, 3.8 laryngitis 1-5 times; OR 4.2 95% CI 2.1, 8.5).

The other health conditions investigated show little consistency between studies. Although all the health conditions that were included in the multiple regression models of studies showed statistically significant associations, there were many health conditions that were not sufficiently correlated with voice problems in the univariable model to be included in the multiple regression. For example, heartburn /reflux were found to have a significantly positive association in studies by

Charn and Mok (2012) and Sampaio et al. (2012), but after stepwise regression Gadepalli et al. (2019) found no association. Asthma was found to be positively associated with voice problems by Lee et al. (2010), but after stepwise regression was not found to show statistical significance by Gadepalli et al. (2019). Some health conditions were found to be statistically significantly associated with voice problems but were only investigated in one study. These include rhinitis (Sampaio et al., 2012), cough (Gadepalli et al., 2019) and gastritis (Assuncao et al., 2012). Without other studies investigating these health conditions, comparisons cannot be made with other teaching populations and therefore conclusions about the associations of these health conditions is not possible.

A major difficulty with coming to conclusions about any of the results is that each study uses different variables in their multiple regression models so that it is difficult to compare results. The only variables which show consistently positive associations are vocal effort, perceived noise and respiratory infections. Other variables show lack of consistency in results and so no conclusions can be made about their associations. It may be that the combination of factors included in each study's multiple regression model may affect the results, and thus different combinations of variables in a model may account for differences in associations. It may be that there are extraneous factors in the teaching population not accounted for that may cause differences in results, this could be a result of a bias in the participants or inherent in the population. Furthermore, limitations in some of the study's designs may have led to incorrect results.

2.4.3 Methodological limitations of studies

A general limitation of all these studies, which all use a cross sectional design, is that it is not possible to determine the direction of associations between risk factors and voice problems. Longitudinal studies would need to be undertaken in order to have more clarity as to the direction of the association. Additionally, it is not possible to infer causality and therefore, results of any of these studies need to be interpreted with this in mind. The methodologies of many of the studies in this review have limitations which need to be considered when interpreting their results. As has been discussed in section 2.4.1, poor reporting, selection of participants, classification of voice problems, response rates and statistical design may have led to biases in the results of the studies. These and other limitations to the studies will be discussed further in this section.

2.4.3.1 Poor Reporting of studies

Poor reporting is a problem with some of the selected studies, thus makes it difficult for the reader to properly analyse the paper and assess its strengths and limitations and allow them to determine whether the results are reliable and valid (Vandenbroucke et al. 2014). Some of the studies do not clearly define their target population, therefore it is not possible to determine the applicability of the results. For example, Assuncao et al. (2012) do not report any eligibility criteria and whether

their participants taught at preschool, elementary or middle school level. So, for instance it is not known whether temporary teachers, part-time teachers or teachers of special education participated. Including particular types of teachers is likely to affect the results and therefore needs to be known. Charn and Mok (2012) also do not adequately report their eligibility criteria. Although they report they recruited teachers from primary schools it is not known whether they were full and part-time, temporary and permanent or included special school teachers.

Of particular concern in some of the studies are discrepancies between the reporting of the results in the description and the tables. This occurs in studies by da Rocha and Souza (2013) and de Medeiros et al. (2008). da Rocha and Souza (2013) report in the text of their multivariate analysis that workloads of more than 40 hours, teaching in rural areas and refraining from vocal rest were significantly related to higher VHI scores, but are shown to be insignificant in the table of results. de Medeiros et al. (2008) report in the description for their univariate analysis that water intake and other activities with intense use of voice were statistically significantly associated with possible and probable dysphonia but the results table shows a non-significant relationship. Therefore, in these studies we cannot be certain of the validity of the results.

2.4.3.2 Statistics

The statistical models for the majority of the studies in the literature review are not appropriate for the study design. None of the studies used a multilevel model in their analysis and thus most do not account for the possibility of clustering in teachers from the same school and thus school effects. Only the study by Akinbode et al. (2014) accounts for this by using a fixed effects logistic model. There may be unobserved characteristics within particular schools that correlate with the outcome measure that need to be accounted for (Rasbash, 2009). From the number of schools involved in many of the studies it seems possible that clustering is possible. For example, Lee et al. (2010) includes 498 teachers from 20 schools, Charn and Mok (2012) 214 teachers from six schools, da Rocha et al. (2015) 573 teachers at 31 schools, de Medeiros et al. (2008) 2103 teachers at 83 schools, Sampaio et al. (2012) 4496 teachers at 365 schools, and Gadepalli et al. (2019) 454 teachers from 44 schools. If clustering is ignored statistically significant associations may be concluded when they do not exist (Buxton, 2008). Furthermore, standard errors are biased towards smaller values with resulting p values that show greater statistical significance than their true value and confidence intervals that are narrower than they really are (Kirkwood and Sterne, 2003).

There are also limitations with the way some of the studies chose the variables for the final model. Four of the studies used stepwise regression (de Alvear et al., 2011; Gadepalli et al., 2019; Lee et al., 2010; Rossi-Barbosa et al., 2016). Many authors have argued that stepwise regression is not an appropriate method for selecting a multiple regression model as it can cause an overestimation of

the results, with confidence intervals being too narrow and p values lower than they should be, and can prevent the 'true' variables from being selected (Flom and Cassell, 2007; Harrell Jr, 2015; Kirkwood and Sterne, 2003; Smith, 2018). Seven of the other studies have p value thresholds, so that only variables that have p values equal or below the threshold are included in the multiple regression. Assuncao et al. (2012), de Medeiros et al. (2008), Rossi-Barbosa et al. (2016), Sampaio et al. (2012) and the Da Rocha studies use a threshold of 0.20 and Moy et al. (2015) uses 0.25. However using thresholds does not account for the fact that variables may only become important in the multiple variable model (Chowdhury and Turin, 2020; Hosmer Jr et al., 2013). This means that these studies may have missed potentially important variables form the final model and give unrealistic results. A full model approach where all variables are entered into the model has been recommended as this ensures that p values are correct (Royston et al., 2009).

2.4.3.3 Recruitment of Participants

There are limitations with the recruitment of taechers in the studies. Akinbode et al. (2014) and Charn and Mok (2012) use a convenience method to recruit schools for their studies. Convenience sampling means that not all schools in the area of the teaching population of interest will have an equal chance of being selected. The schools selected and thereby the teachers within those schools, may not be representative of the population of interest. Teachers may be biased in terms of specific characteristics such as age, gender, ethnicity, voice use etc. Therefore, the external validity in these studies may be compromised and the results biased. Charn and Mok (2012) go some way into checking the representativeness of their population by comparing gender and age to Singapore's national teacher statistics, finding similar proportions. However, they did not account for other demographic characteristics such as ethnicity, so there may be sub-groups that are not representative of the larger teaching population.

Da Rocha and Souza (2013), da Rocha et al. (2015), Lee et al. (2010) and Moy et al. (2015) randomly select schools for their studies so schools are more likely to be representative of all schools in the area of interest. However, selecting schools randomly rather than carrying out random sampling of teachers is more likely to cause sampling error and therefore the participants are less likely to be representative using this method (Ruane, 2016). Furthermore, there is likely to be cluster effects within the schools, so that teachers in the same school are likely to be more similar to each other than teachers in other schools in terms of certain variables, due to certain factors within the school.

.

2.4.3.4 Classification of Voice Problems

A particular limitation in many of these studies is the classification of voice problems and the resulting potential for measurement error. Misclassification of an outcome variable may produce incorrect associations between outcome and exposure variables (Vandenbroucke et al., 2014). This is likely to occur in the studies not using tools that have been assessed for reliability or validity (Assuncao et al., 2012; Akinbode et al., 2014; Charn and Mok, 2012; de Alvear et al., 2011; De Medeiros et al., 2008; Lee et al., 2010; Rossi-Barbosa et al., 2016)

For example, Rossi-Barbosa et al. (2016) classify a voice disorder as a participant responding 'yes' to the question "Have you noticed changes in your voice quality?" The limitation with this statement is that voice quality is not defined in the study and participants may interpret it differently to each other, especially in terms of severity. For example, some participants may have not noticed changes in their voice quality but may be having difficulties with their voice in terms of vocal fatigue or finding it difficult to speak. Furthermore, a participant's voice quality could have improved, for example, after having treatment for voice problems, voice coaching or singing lessons and therefore, responding 'yes' to this question would misclassify them as having a voice disorder.

De Alvear et al. (2011) classify a voice problem as the simultaneous presence of vocal effort and two or three vocal symptoms. The study does not define vocal effort so that participants may interpret it differently. Additionally, those who already have a voice problem may use a microphone to help prevent them from straining their voice and therefore they may respond 'no' to this question. Therefore, they would then be incorrectly misclassified as not having a voice problem. This study only gives a list of three possible voice symptoms for participants to choose from: (i) dry, sticky, aching or burning throat (ii) feels like keeping silence because voice is fatigued (iii) voice timbre has changed and become hoarser. However, there is a possibility that a participant has other symptoms that may indicate they have a voice problem, for example, repetitive throat clearing, sensation of lump in the throat and voice interfering with communication.

Although using validated questionnaires is more likely to correctly classify voice problems, as discussed in detail in section 2.4.1.4, there are problems with the way that some of the studies employ these tools so misclassification may still have occurred.

2.4.3.4 Questionnaire pretesting

Piloting or pretesting a study is important when carrying out a survey design in order to ensure that the questionnaire is fit for purpose. Pretesting allows the researcher to check questions are understood as intended (Hilton, 2017) and the response options are appropriate (Ruel et al., 2015).

If questionnaires do not get pretested, there is a possibility that questions may be misinterpreted or not suitable, causing measurement errors (Creswell and Creswell, 2018).

In the review only two of the studies give sufficient information about their pretests that allows the reader to assess its purpose and outcome (Akinbode et al., 2014; Gadepalli et al., 2019). Some of the studies do report pretesting their questionnaire or carrying out a pilot study but give limited or no information about the details (da Rocha and Souza, 2013; da Rocha et al., 2015; Moy et al., 2015; Rossi-Barbosa et al., 2016). The remainder of the studies do not report pretesting their questionnaire at all (Assuncao et al., 2012; Charn and Mok, 2012; De Alvear et al., 2010; De Medeiros et al., 2008; Lee et al., 2010; Sampaio et al., 2012). Therefore, for most of the studies it is not possible to assess whether the participants were likely to interpret the questions correctly. This lack of information compromises the validity of the study as incorrect results due to participants misinterpreting questions may be a possibility.

2.5 Conclusions

This literature review shows that although there has been a lot of research into risk factor studies there is little conclusive evidence that links specific factors with voice problems. The results from studies done with human cell materials and animal larynges are not ecologically valid. Although these studies provide interesting and potentially relevant results, studies with human participants are needed in order to provide results that will be relevant to a clinical population or those wanting to prevent voice problems. Experimental designs likewise lack ecological validity and can therefore only theorise about links between risk factors and voice problems. Thus, the associations between risk factors and voice problems need to be observed in real life situations.

A literature review was carried out of observational studies with teachers. There was variability in the methodologies of the studies making them difficult to compare. This included different risk factors investigated, teaching populations and the way the outcome was measured. Greater similarities over the risk factors included, and the outcome measures used would make it easier to come to more definite conclusions on the associations between risk factors with voice problems.

There was a lack of quality in the methodology of some of the studies with only half using validated tools to measure voice problems (da Rocha and Souza, 2013; da Rocha et al., 2015; da Rocha et al., 2017; Gadepalli et al., 2019; Moy et al., 2015; Sampaio et al., 2012). Although many studies selected schools randomly, others use convenience sampling (Akinbode et al., 2014; Charn and Mok 2012) which may have detrimentally effected representativeness and generalisability. Only one of the studies accounted for the clustering of teachers from the same school in their statistical design (Akinbode et al., 2012). As Galbraith et al. (2010) suggests, clustering can significantly impact on

data analysis and a study's conclusions. Furthermore, a large drawback was insufficient reporting of different aspects of each study's methodology meaning that it was difficult to interpret results. Therefore, due to limitations with the methodologies of these studies the validity of the results may be problematic. As a result, there is a need for more good quality studies to fully understand the associations related to voice problems. Although there were consistencies in the associations between some risk factors and voice problems, in particular, vocal effort, noise, psychological factors and upper airway infections/problems, there was variability in the results across studies for many risk factors. These included gender, age, lifestyle factors, health and environmental factors. These inconsistencies have implications on the management of patients with voice problems. Greater clarity is needed around the risk factors that are associated with voice problems for successful treatment and prevention.

The review has also demonstrated that although there have been studies investigating the association between psychological factors and voice problems (Assuncao et al., 2012; da Rocha and Souza, 2013; da Rocha et al., 2015; da Rocha et al., 2017; de Medeiros et al., 2008; Gadepalli et al., 2019; Moy et al., 2015), all of these were investigating negative psychological factors and has therefore highlighted the need for studies to investigate the association between positive psychological factors such as wellbeing and voice problems.

With the limitations to the studies in the review and the consequential lack of conclusive evidence for associations between risk factors and voice problems and lack of research into wellbeing, a nationwide study carried out in the UK including wellbeing and other risk factors would be a useful addition to the literature.

Chapter 3. Methodology

3.0 Introduction

This chapter provides information on the methodological process relating to this research. It begins by discussing the research design for the study and the rationale for using a cross-sectional design. This is then followed by discussing the sampling method and outlining the eligibility and exclusion criteria for participants. Following this, the chapter reports on the data collection procedures, and the pretesting of the survey. Ethical considerations for the study are then explored, including informed consent, confidentiality, diversity and equality, and the impact of the research. This is then followed by considering the design of the questionnaire looking in detail at the published measurement tools used within the study, the Voice Symptom Scale (VoiSS) (Deary et al., 2003), Warwick Edinburgh Mental Wellbeing Scale (WEMWBS) (Tennant et al., 2007) and the GerdQ (Jones et al., 2009). Finally, the procedures for the statistical analysis are outlined. The primary purpose of the study is to see if there is a relationship between wellbeing as measured by scores on the WEMWBS and voice symptoms as measured by the total score on the VoiSS.

3.1 Research Design

The design was an epidemiological, cross sectional study to assess the association of wellbeing and other risk factors with voice symptoms, as measured by the total score on the VoiSS (Deary et al., 2003) in teachers working in England. Using a cross-sectional design allows the measurement of voice problems and exposure status at one point in time (Rothman et al., 2008). All participants in the target population regardless of the severity of their voice problem or exposure status have an opportunity to participate. This therefore may reduce the bias in participants taking part in the study. For example, it allows for those with voice problems that may not have been formally diagnosed to be included, and therefore a more representative population of teachers with voice problems may participate. A cross-sectional study is an efficient and effective way to gain information from large numbers of participants. However, before committing to this design other observational designs were considered. These included case-control studies and longitudinal studies.

In a case-control study, participants are selected according to the disease they have and are then compared against those without the disease (Kirkwood and Sterne, 2003). However, using this design could encounter difficulties in recruiting an unbiased sample. In order to employ this design, it would have been necessary to recruit teachers with voice problems separately from those without voice problems. There were two possible approaches to achieving this. The first possibility could

have been to contact schools and ask teachers with voice problems to take part in the study. However, as discussed in the literature review, defining what a voice problem is, is not easy to establish and therefore misdiagnosis could occur.

The second approach would be to ask teachers with a formal voice disorder diagnosis to take part in the study. However, it is likely that a significant number of people with voice problems do not approach their doctors. This would mean that the sample would be biased towards the type of people who go to their doctors with voice problems. A review carried out by Campbell and Roland (1996) looking into why people consult their doctors, suggested that those who consulted their doctors were more likely to have conditions that were increasing in severity, those who thought stress was associated with their illness, those who wanted more knowledge about their condition and those living close to their GP surgery. A cross-sectional design could potentially have less bias in selecting participants with voice problems, as it will be inviting all teachers to take part whether or not they think they have a voice problem. It is therefore likely to recruit all types of teachers with voice problems including those who may not consult a doctor.

A longitudinal study was also considered. Although both cross-sectional designs and longitudinal designs can collect information on many variables a longitudinal study has the advantage over a cross-sectional study as there is more certainty over the direction of an association (Gordis, 2014). For example in a cross-sectional study if there is an association between poor psychological health and voice problems it will be unclear which came first, as a cross-sectional study only surveys people at one point in time, whereas in longitudinal studies measurements are taken over time. However, longitudinal studies require sufficient time and resources to enable repeated testing. Due to time and resource constraints, it was only possible to investigate the population at one point in time. Therefore, a cross-sectional design was selected as the most appropriate design.

3.2 Questionnaire design

3.2.1 Introduction

A self-reporting questionnaire was developed on the web platform Bristol Online Survey (BOS). A full copy of the questionnaire can be seen in Appendix 4. The questionnaire was designed according to research in the literature on voice problems, and was also informed by large health questionnaires used in the UK, such as the 1958 National Child Development Study (University of London, 2015), 1970 British Cohort Study (University of London, 2016), English Housing Survey 2014 (Department for Communities and Local Government, 2017) and Health Survey England 2015 (NatCen Social Research, 2019).

The questionnaire contained items regarding socio-demographics, the teaching environment, vocal behaviour, health and lifestyle, and wellbeing.

Variables that were included in the questionnaire and measured included:

- Voice symptoms measured by total scores on the VoiSS (Deary et al., 2003)
- Wellbeing measured by total scores on the WEMWBS (Tennant et al., 2007b)
- Gastroesophageal disease as measured by total scores on the GerdQ (Jones et al., 2009)
- Number of years teaching
- The average number of hours teaching in a week
- Age in years
- Classroom size (average number of pupils)
- Sex (male/female/prefer not to say)
- Asthma diagnosis (no/yes)
- Respiratory infection in the last 30 days (no/yes confirmed by a doctor/yes not confirmed by a doctor)
- Talking louder than normal in the classroom in the last 30 days (always/most of the time/sometimes/never)
- Talking over background noise in the last 30 days (always/most of the time/sometimes/never).
- Subject taught
- Voice information or training (no/yes teacher training/yes professional development/ yes teacher training and professional development)
- Smoking status (every day/never/occasionally/used to)

At the beginning of the questionnaire, participants were asked to give details about their schools including the name and address of the school, the type of school (categories: mainstream, special school or alternative provision/PRU), and whether it was a state school or fee-paying school (categories: Community school, Academy, Free school, Voluntary aided, Voluntary controlled, Foundation, Private fee-paying school). They were also asked to provide details about their ethnicity. After data collection, information about the type of school was checked against available data collected by the Department for Education. The region of the school, the area of England where the school was situated, was also added to the data after data collection had taken place. The regions were West Midlands, East Midlands, East of England, Yorkshire and Humber, North West, North East, London, South West and South East.

3.2.2 Outcome Variable

Voice problems were measured using the Voice Outcome Scale (VoiSS) (Deary et al., 2003). When answering questions on the VoiSS participants were asked to consider how their voice had been over the last thirty days. The participants were required to fill in all questions on the VoiSS before they could proceed with the rest of the questionnaire. This was because the analysis was reliant on a total score for the VoiSS. If any of the participants had not filled in every question on the VoiSS, their data would have to have been removed. The total VoiSS scores for each participant were calculated by the researcher once data collection had been completed.

The VoiSS is a 30 item self-administered questionnaire that measures how someone perceives their voice. Wilson et al. (2004) used principal components analysis to reveal a three-factor structure of impairment, emotional response and physical symptoms. Questions in the impairment domain are those related to day-to-day functioning that may be impacted by the voice and voice quality. Examples of questions include: 'Is your voice hoarse?' and 'do you have difficulties attracting attention?' The physical domain includes questions that ask about physical problems that may affect the voice, for example, 'Is your throat sore?' and 'does it feel as if there is something stuck in your throat?' The emotional domain is related to feelings that are affected by the voice problem, for example, 'do you feel miserable or depressed because of your voice problem?' and 'are you embarrassed by your voice problem?'

The VoiSS uses a frequency scale for its scoring system. Deary et al. (2003) piloted a five-point frequency scale (all the time, most of the time, some of the time, occasionally, never) and a severity scale which considers how bad a participant feels their symptom is (unbearable, severe, moderate, slight, not at all). Deary et al. (2003) found participants were more likely to fill in the frequency scale so this was used in subsequent versions of the VoiSS. The VoiSS scores are 0 for 'Never', 1 for 'occasionally', 2 for 'some of the time', 3 for 'most of the time', and 4 for 'always'. The scores for individual items are added together to get a total score. The total possible range of scores for the VoiSS is 0 - 120, with higher scores related to poorer voice functioning.

Before considering using the VoiSS as the tool to measure the outcome variable, several self-assessment tools for voice problems were evaluated using the Scientific Advisory Committee of the Medical Outcomes Trust (2002) as a framework. Other voice Related Patient Report Measures that were considered and evaluated were the Voice Handicap Index (VHI) (Jacobson et al., 1997), Voice Handicap Index – 10 (VHI-10) (Rosen et al., 2004) and the Vocal Performance Questionnaire (VPQ)(Carding et al., 1999). The Voice Related Quality of Life Measure (V-RQOL) (Hogikyan and

Sethuraman, 1999) and the Voice Activity and Participation Profile (VAPP) (Ma and Yiu, 2001) were also initially considered but neither had been validated with UK populations so would not be appropriate to use with teachers in England.

The development of each instrument and the results from validation studies were compared between the VoiSS, VHI, VHI-10 and VPQ. The VoiSS and VPQ were developed in the UK and therefore may be more valid to use with teachers in England than the VHI, which was developed in the US. The VPQ however does not outline its development process and therefore it is not possible to determine its validity. The Scientific Advisory Committee of the Medical Outcomes Trust (2002) states a patient-reported outcome scale should include its development process as it is an important element to assess validity.

The development of the VoiSS involved an in-depth process. Initially 133 voice patients were asked to list any difficulties they had related to their voice problems (Scott et al., 1997). From these, 53 useful items were identified and included in a pilot questionnaire. In the next phase 168 participants with a variety of voice problems, completed the pilot questionnaire and using Principal Components Analysis (PCA) the items were reduced to 31. Thirteen items from the VHI were then added to give a 44 item questionnaire and one item regarding employment was removed. The resulting 43 item questionnaire was completed by 319 voice patients and further reduced to 30 items using PCA. A detailed discussion of this analysis can be found in Deary et al. (2003) and Wilson et al. (2004)

The VoiSS shows good reliability and validity in UK populations (Webb et al., 2007). Cronbach's alpha was 0.89 for the total score, and test-retest scores tested 1 week apart, were 0.63(0.43-0.79) for the total score (Webb et al., 2007). Although the VHI and VPQ had higher test-retest scores with the same population, 0.83 for the VHI and 0.75 for the VPQ, the researcher decided to use the VoiSS because of its robust development process demonstrating good validity.

For this study, the VoiSS was slightly adapted for the population, so that all teachers including those that did not regard themselves as having a voice problem would fill it in. Therefore, any question that included the words 'voice problem' were slightly adapted with the word 'problem' taken out of the question. So, for example the question 'are you embarrassed by your voice problem?' became 'are you embarrassed by your voice?' This related to questions 10,13,18,29 and 30.

3.2.3 Continuous Independent Variables.

3.2.3.1 Mental Wellbeing

The assessment of wellbeing was achieved using the Warwick Edinburgh Mental Wellbeing Scale (WEMWBS), which consists of 14 items (Tennant et al., 2007). Participants are asked to consider each statement on the questionnaire in relation to the past 2 weeks. The WEMWBS has a 5-point

frequency scale ('none of the time', 'rarely', 'some of the time', 'often', 'all of the time'). A score of 1 is given to 'none of the time' and 5 to 'all of the time' with a possible range of scores of 14 to 70 and higher scores indicating higher wellbeing. The researcher calculated the total scores for the instrument once data collection had been completed.

The questionnaire was chosen as it was developed in the UK with a general population and is therefore likely to be a valid measure to use in this study. It has been psychometrically tested in British populations and is used in many large UK health surveys, for example, Health Survey for England (NatCen Social Research, 2015). Good internal consistency has been established in large randomly selected general Scottish and English populations and therefore is likely to be generalisable to other UK populations. Tennant et al. (2007) calculated a Cronbach's Alpha of 0.91 using data from the Scottish Health Education Population Survey (Gosling et al., 2007) and "Well? What do you think?" Survey (Braunholtz et al., 2007) and Fat et al. (2017) calculated a Cronbach's alpha of 0.92 using data from Health Survey for England.

The shorter seven-question version of the WEMWBS, the SWEMWBS was also considered. It has a lower participant burden, which may make it more likely to be completed in a survey than the WEMWBS. However, the WEMWBS showed a better internal consistency score than the SWEMWBS which was found to be 0.84 (Fat et al., 2017). Furthermore, data for the SWEMWBS was extracted from the WEMWBS (Fat et al., 2017). The results for the SWEMWBS may not be reliable, as participants may have answered differently if they had been given the SWEMWBS separately.

3.2.3.2 Gastroesophageal Disease

Gastroesophageal disease (GERD) was measured using the GerdQ questionnaire which has been designed to help clinicians diagnose and manage GERD (Jones et al., 2009). The total possible scores are 0-18 with increasing scores indicating the increasing likelihood of GERD. A cutoff score of 8 with a specificity of 71.4% and a sensitivity of 64.6% indicates a high likelihood of having GERD. Scores below 8 indicate a low or no likelihood of having GERD (Jones et al., 2009). However, there is no gold standard for diagnosing GERD so the GerdQ was measured against the best diagnostic practice using endoscopy, wireless 48h pH recording and Symptom Association Probability (SAP). Therefore, the specificity and sensitivity results need to be interpreted with caution. Because of this, the researcher decided not to use the cut off scores but use the GerdQ as a continuous measure. The questionnaire comprises of six questions. Four of these questions relate to positive predictors of GERD; heartburn, regurgitation, sleep disturbance, and use of over the counter reflux medication. Two of the questions relate to negative predictors of GERD, epigastric pain and nausea. Participants filling in the scale are asked to reflect on their symptoms over the previous seven days. For the positive predictors, a score of 0 relates to 0 days, a score of 1 to 1 day, a score of 2 relates to 2 to 3

days and a score of 3 relates to the symptoms occurring 3 to 7 days over the last seven days. For the negative predictors, scores are reversed, so a score of 0 relates to symptoms occurring 3-7 days over the last 7 days, etc. The researcher calculated the total scores for the instrument after data collection had been completed.

3.2.3.3 Deprivation Score of the School

The deprivation of the school was measured using Pupil Premium. Pupil premium refers to extra money a state-funded school is given to help disadvantaged children improve their academic performance. This includes children who receive free school meals, and looked after and previously looked after children (Department for Education, 2020). Participants were not required to answer questions about pupil premium. The pupil premium figures for 2017 to 2018 were extracted from the Department for Education's 'Pupil premium 2017 to 2018: allocations (school level)' Excel spreadsheet (Department for Education, 2018a), into the Excel spreadsheet containing the research data, after data collection had taken place. The numbers for each school are recorded in percentages and relate to the percentage of children for which extra funding has been given.

3.2.3.4 Other continuous variables

Other variables that were included were Age (in years), Hours teaching and Number of years teaching. All variables were included as the current evidence indicates a possible link with voice problems. For each of these variables, participants were asked to select a number from a drop-down menu. Although age was being treated as a continuous variable in the analysis, it was also converted into categories in order to be able to compare the age of the study's population with that of the English population as it appears in the School Workforce Census (Department for Education, 2018b). This was done to calculate bias and assess representativeness of the study population. The age categories were: 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60 and over.

3.2.4 Categorical Variables

The categorical variables that were included were vocal training, talking louder than normal, talking above background noise, asthma diagnosis, smoking, sex, respiratory infection, subject taught and class size. All were included as there is evidence to suggest there may be links with voice problems as shown in the literature review.

For the variables talking louder than normal and talking above background noise, a four-point frequency scale was used. These were 'never', 'sometimes', 'most of the time', 'always'. Questions were devised that attempted to find objective ways to ask about vocal loading and were informed by questions used in other risk factor research for teachers. Although there are difficulties with Likert scales, e.g., they presume equity of intervals and individuals may have different interpretations of

the response categories (Jones and Loe, 2013), they are simple to use. Asking respondents to give answers in a numeric discrete format, i.e. number of hours would have been a more objective measure but would have increased participant burden and may have led to measurement errors due to difficulties in recall (Groves, 2009).

For the variable 'subject taught', for ease of analysis, the subjects that were included on the questionnaire were broken down into fewer categories after data collection. The categories Primary classroom and PE were kept as they were. PE was kept as a separate subject as there is evidence to suggest that PE teachers use more vocal loading than teachers from other subjects, and therefore may be more at risk of getting voice problems (Kristiansen et al., 2014). Art and design, English, citizenship, history, design and technology, humanities, languages, leisure and tourism, media studies, PSHE, and RE were put into the category 'arts and humanities'. Biology, chemistry, computing and IT, maths, physics and science were put into the category 'science'. Dance, drama and music were put into the category 'performing arts' and business and economics, geography, politics, law and psychology were put into the category 'social science'.

Class size was collected as a continuous variable. Participants were asked to estimate the typical class size they had taught over the previous 30 days. They were given a drop-down menu from which to select the number. The largest option on the drop-down menu that was given was 'over 40'. Nine people chose this option. Therefore, to ensure their data was used in the analysis the data was put into two categories, 'under 29' and '29 and over'. To decide on the cut-off point for classroom size, previous risk factor research with teachers was reviewed. These included studies by Akinbode et al. (2014), da Rocha et al. (2017) and van Houtte et al. (2012). Akinbode et al. (2014) included the class size categories, 2-24, 25-35 and 36-94, da Rocha et al. (2017) used two categories, 'up to 25 students' and '26 or more students' and a study by van Houtte et al. (2012) carried out in Belgium used 3 categories, <15, 15-20 and 21-25. However, these studies all include primary school teachers only, so may not be generalisable to a study that uses secondary school data. Additionally, as the studies have not been carried out in the UK, they may not be suitable categories for UK teaching populations. Certainly, the categories in the van Houtte et al. (2012) study, which only allows class sizes of up to 25, would not capture the data in this study. Furthermore, none of the authors explains their choice of categories. The researcher looked at other quantitative studies with UK teachers but did not identify any that had used class size as an outcome variable. The researcher therefore dichotomised the data by choosing the median as the cut-off point.

3.3 Sampling Method and Recruitment

Participants were primary and secondary school teachers in England who were invited to complete an online survey. Primary and secondary school teachers from state schools, independent schools, mainstream schools, special schools and Pupil referral units (PRUs)/alternative provision were invited. The aim of the study was that the results of the analysis would be as generalisable to as many teachers in England as possible. The inclusion and exclusion criteria are outlined in the sections below.

3.3.1 Inclusion Criteria

The inclusion criteria were teachers with permanent contracts working in primary or secondary schools who worked with pupils from Reception age up to and including 'A' levels. Teachers could be working in Mainstream, special or alternative provision schools and in state schools or independent fee-paying schools. Schools were identified using a database provided by a school's marketing company. Broad inclusion criteria meant that the results would be able to be generalised from the study sample to a wide population (Kendall, 2003).

3.3.2 Exclusion Criteria

Teachers that were excluded from the study included Nursery and Early Years Teachers who taught pre-reception age, those who taught in further education, and those who were not permanent members of staff such as music peripatetic teachers and supply teachers. These groups were excluded because they may have different teaching patterns to other teachers. For example, Hutchings et al. (2006) reported that supply teachers have different working patterns to the national teaching population in permanent jobs. As well as having no classroom responsibilities Hutchings et al (2006) suggested that on average supply teachers work at six schools a year, have a higher proportion of overseas-trained teachers and often do not teach in a regular pattern.

To ensure that only those who were eligible completed the questionnaire, a question was included on the web-based questionnaire that screened out ineligible participants. It asked whether participants taught at Further Education or Nursery Level or as a supply teacher or peripatetic music teacher. Those who responded 'yes' were screened out of the questionnaire.

A census-based approach was taken in recruiting participants. A census-based approach means that every unit of the population is given the opportunity to take part, so that every school in England was invited to participate in the research. Initially, the objective of the study was to send out a web-based questionnaire to a representative sample of teachers in England using a sampling frame that consisted of a list of members of the teaching profession. However, it was not possible to access a list of members of the teaching profession and therefore contact teachers individually. From 2000

until 2012, the General Teaching Council for England (GTCE) maintained a register of teachers in England. However, the GTCE was decommissioned in 2013 and became the National College for Teaching. The researcher contacted the National College for Teaching to ask them if they could assist in selecting teachers for the research but was informed that this would not be possible.

A clustered design where schools were randomly selected was also considered to help control for bias as much as possible. However, it is acknowledged that response rates to online questions are often very low (Anand et al., 2015; Cunningham et al., 2015; Kirkby et al., 2011). Published response rates are derived from individuals being sent on-line surveys, rather than organisations such as hospitals or schools which are asked to distribute the questionnaires to their employees. Figures are likely to be even lower when individuals are not contacted directly (Eysenbach, 2004), as is the case with this research. Therefore, to get as large a number of completed questionnaires as possible, all schools in England were sent the questionnaire.

A web-based questionnaire was employed as this was the most efficient way of collecting the required information with the time and financial constraints of an MPhil. The benefits of online surveys include shorter delivery time, reduced costs, and fewer measurement errors when compared with other methods (Groves, 2009). Data collected by other methods have to be manually inputted and thus it is more likely to incur errors than using a web-based survey. It was also expected that a web-based survey would create less participant burden for teachers than a paper-based survey where teachers would have to fill in handwritten responses and have to find somewhere to post the questionnaire. Teachers as part of their jobs use computers so accessibility to the internet should not pose a problem. The survey was delivered using the Bristol Online Survey (BOS), now called 'Online Surveys'. BOS was used as Birmingham City University (BCU) has a licence for its use and it returns data that is suitable for further analysis, for example, Excel and CSV.

Data collection took place between 7th November 2017 and February 18th, 2018. Initially, all schools in England were contacted via email with a link to an online web-based self-administered questionnaire, inviting them to participate in the study (Appendix 5). The email requested that the link to the questionnaire be forwarded to the school's teaching staff. The survey was not password protected so that any teacher would be able to open the survey. A reminder email was sent on 6th December 2017.

Due to the initial low number of responses (258 between 7th November and 6th December), a more targeted approach to recruiting participants was devised (Lynn, 2017). The following methods were employed:

1. The researcher approached BCU School of Education who provided a list of schools in the West Midlands, for which they had specific contacts, including head teachers and teaching placement mentors. These contacts were approached directly via email. The emails were addressed personally to the head teacher or teaching mentor.

2. Teaching unions NASUWT, NEU and Voice were contacted via email asking them to distribute details of the study to their members. The Voice agreed to include the details of the study in their newsletter.

3. The researcher used Twitter and Facebook as an engagement tool to encourage users to fill in the questionnaire. Due to the popularity of these platforms, it was hoped that a significant number of teachers would be made aware of the study. The researcher tweeted teaching organisations with a link to the questionnaire, asking them to retweet to their followers. Messages were also sent to organisations on Facebook asking them to post a link to the questionnaire on their page. It was acknowledged that social media platforms are not representative of populations (Ruths and Pfeffer, 2014; Sloan et al., 2015). However, as the initial mailshot had been unbiased in its delivery method it was decided subsequent approaches to data collection were appropriate in order to get a sufficient number of respondents. There are ethical issues regarding engaging users on social media platforms concerning privacy and confidentiality that needed to be considered (Jones, 2011). However, as Twitter and Facebook were only being used as a tool to promote the study and not being used to access data that users had published, it was felt that this would not be an issue.

4. Personalised emails were sent to head teachers and teachers in other parts of the country as well as the West Midlands. School email addresses were accessed from the website 'schools web directory'. As there was not enough time to send out personal emails to all head teachers in the country a random method of selection was devised. As the schools were arranged by county, counties were selected randomly. Every school in the county for which email addresses for head teachers or teachers could be acquired were approached. In total 1205, teachers from 648 schools took part.

3.4 Non-Response Bias

Non-response bias is a possibility for web-based surveys especially as response rates are generally low (Fricker et al., 2002). The literature suggests that those who are more interested in the topic area being investigated may be more likely to respond (Dillman et al., 2014). Therefore, this study may be more biased towards people interested in voice problems, perhaps those who have had a voice problem in the past or who have a current voice problem. To try to account for this, it was stated in the covering email the researcher would like to hear from both teachers who had voice

problems and those who had never had voice problems. Non-response is very difficult to assess, as non-respondents were not given the opportunity to say why they had chosen not to participate. Furthermore, it was not possible to determine which teachers in the population chose not to respond to the questionnaire and those who did not receive the questionnaire. However, research has identified that certain groups may be less likely to respond to questionnaires. These include men and young people (Roberts et al., 2004) and those from black and ethnic minority groups (BAME) (Care Quality Commission, 2020; Sheldon et al., 2007). Some of this is reflected in the results section, which measured the bias of participants that had completed the survey by comparing the characteristics of participants in the study with those of the teaching workforce in England. For example, it found that men and teachers from some BAME groups were less likely to participate in the study. This will be discussed in further detail in results chapter.

3.5 Pretesting the Questionnaire.

As part of the development of the questionnaire, the survey was pretested with a subpopulation of teachers. These included eight teachers across three schools in Birmingham, two primary schools and one secondary school. A convenience sample was obtained through contacts of the researcher. The pretesting took place in September 2017. Six of the participants completed a questionnaire with the researcher present. During the process, the participants were given the opportunity to provide comments about the questionnaire and suggest amendments. Two of the participants completed the survey remotely and emailed the researcher any comments or suggestions after completion. Research latency was also assessed. This is the amount of time it takes to complete individual items on a questionnaire and to complete the full survey (Ruel et al., 2015). The length of time it took to complete the survey was reported in the participant information at the beginning of the survey. As a result of the pretest, a few minor amendments were made to the questionnaire. The schools involved in the pretesting were not invited to participate in the wider study. Involving participants that have been involved in the pre-test may cause bias. Those who have taken part in the pre-test may respond differently to those who have not previously been involved in the study (Van Teijlingen and Hundley, 2010).

3.6 Ethical Considerations

It is the researcher's responsibility to ensure that research takes into account any ethical issues that may be encountered before data collection takes place (Gray, 2013). The researcher applied for ethical approval from the ethics committee from the Faculty of Health, Education and Life Sciences (HELS) at BCU by completing and submitting all relevant paperwork. This included providing information about the ethical dimension of the project and supporting documentation including a

research proposal, participant information and consent form. Ethical approval was given on 15th August 2017. Appendix 4 shows the letter for ethical approval.

For ethical approval to be granted the following considerations were made:

3.6.1 Informed consent

To ensure informed consent, participant information was included at the beginning of the questionnaire. The participation information clearly explained all aspects of the study including the aims of the project, making it clear that participation was voluntary. It also stated the benefits and potential risks of the project, how the data would be stored and for how long, and contact details for those involved in the project. Consent was obtained on the questionnaire itself. Participants were asked to read a list of statements and indicate whether they agreed with them. This was achieved by providing participants the option to choose 'yes' or 'no'. Participants who selected 'yes' had given their consent and could proceed to the next question on the questionnaire. Those who selected 'no' did not consent and were screened out of the questionnaire. Contact details were given so that a participant could clarify any of the information included in the participation information or request further information about the project. Participants were also informed that even if they had completed a questionnaire, they could contact the researcher at any time during the data collection period to have their entry removed. In order for participants to be able to do this, on completion of the questionnaire a reference number was provided to each participant. Appendix 1 shows the participant information and Appendix 2 shows the consent to participate.

3.6.2 Confidentiality

It is important that the confidentiality of participants involved in research be maintained (WMA, 2001). As the individual names of teachers were not collected, the data was confidential. However, teachers were asked to provide the names and addresses of the school(s) they taught in which may have caused some teachers to be identifiable and therefore could not be considered anonymous. For example, as the data being collected were very comprehensive, certain details given by an individual collated together may have identified them. Therefore, to minimise the risk of identification, each school was given a code generated randomly in Excel. Participants were assured that names of schools would not be included in any publications, conference proceedings or data sharing. At the time of the ethics application, the storage of data complied with the Data Protection Act (1998), i.e. the raw data would only be used for the purposes of the research project and only be accessible to the researcher and supervisory team, stored in a secure location on a password-protected, encrypted computer. However subsequent to the ethical approval, on 25th May 2018, the General Data Protection Regulation (GDPR) came into effect, which superseded the Data Protection Act (1998). Although there are many similarities between the two, the GDPR ensures

more transparency and accountability from organisations over the security of personal data (ICO, 2018). To ensure that the research complied with the GDPR the researcher had an interview with the GDPR lead at BCU and carried out online training on Information Security provided by BCU.

3.6.3 Diversity and Equality

All members of the population should have an equal opportunity to participate in research and research should be accessible to all members of the relevant population (Emanuel et al., 2000). This includes making sure that research is not discriminatory in terms of protected characteristics which the Equality Act (2010) states as being age, disability, gender, reassignment, race, religion or belief, sex, sexual orientation, marriage, civil partnership and maternity. Therefore, all teachers in England were given the opportunity to take part in the questionnaire. Due to the nature of their job, teachers were unlikely to have a problem accessing and understanding any aspects of the questionnaire including participant information and consent. Their job presumably means that they have regular access to the web. However, there may be those that have more difficulty with accessing the survey due to specific learning disabilities, e.g. dyslexia or physical difficulties. BOS takes into account those that may have disabilities with the design of the surveys. These include allowing participants to change the size of the font and colour scheme and do not require the use of the mouse to complete the form.

3.6.4 Impact of Research

The benefits of any research should be maximised whilst any risks and potential for harm are minimised (Beauchamp and Childress, 2013). However, there are inherent challenges in conducting web-based research related to harm of participants. Although the majority of online research poses little risk to participants, any risk needs to be considered as the researcher is not present to protect a participant (Holmes, 2009). Although the risk was deemed small, some participants may have been concerned about their vocal health or other health conditions after completing the questionnaire or been affected by potentially sensitive questions, such as those related to well-being. To address possible problems the participant information advised participants who were concerned about vocal or health issues to contact their GP. Although participants were informed that there would be no immediate benefits to taking part in the project, they were told that the long-term benefits of the research would be to provide information on the factors that may be associated with getting voice problems. This would be useful information for healthcare professionals in treating or helping to prevent voice problems, and for the teaching profession. As well as being an ethical necessity, specifying the benefits of research can help to increase survey responses (Dillman et al., 2014).

3.7 Statistics

3.7.1 Introduction

The primary objective of the analysis was to determine whether there was an association between voice symptoms as measured by total scores on the VoiSS (Deary et al., 2003) and mental well-being as measured by scores on the WEMWBS (Tennant et al., 2007). The secondary objective of the analysis was to see if the other variables of interest were associated with total scores on the VoiSS.

The continuous variables that were included for analysis were:

- deprivation of schools as measured by pupil premium
- age in years
- the average number of hours taught in a week
- number of years being a teacher
- likelihood of GERD as measured by the GerdQ

The categorical variables were:

- sex
- asthma diagnosis
- respiratory infection in the last 30 days
- smoking status
- class size
- talking louder than normal in the classroom over the last 30 days
- talking against background noise in the classroom in the last 30 days
- vocal training or information on how to care for the voice

A multilevel model was carried out to take into account the possible clustering effects of teachers in the same school.

All statistical analyses were implemented using the statistical package R. In each of the tests, statistical significance was determined at a critical value of 0.05 and reported alongside 95% confidence intervals. A p-value assesses the strength of the evidence against the null hypothesis with a smaller p-value indicating stronger evidence (Kirkwood and Sterne, 2003). The 95% confidence interval says that in 95% of repeated samples the population value of the test statistics would lie in the 95% confidence interval window (Kirkwood and Sterne, 2003). Therefore, it allows the reader to see what the likely range of values for a coefficient would be. The confidence interval of 95% is arbitrary and other levels such as 90% and 99% are also used (Hazra, 2017; Tan and Tan, 2010). However, 95% is the most commonly used confidence interval in health research and therefore provides a useful baseline from which to compare other studies.

3.7.2 Null and alternative hypotheses

A null and alternative outcome hypothesis was created for the analysis of the total scores of the WEMWBS against total scores on the VoiSS.

The null hypothesis: There is no association between voice problems, as measured by total scores on the VoiSS, and mental wellbeing, as measured by total scores on the WEMWBS

The alternative hypothesis: There is an association between voice problems, as measured by total scores on the VoiSS, and mental wellbeing, as measured by total scores on the WEMWBS

3.7.2.1 Excel

The data was downloaded from BOS in a zero-indexed format and saved as a CSV file. Zero indexed format meant that all coded variables that were downloaded started at the number 0. Columns that did not include data, for example, participant introductions to the VoiSS, WEMWBS and GerdQ were deleted from the data file and saved as a separate file. As data was potentially identifiable, numerical coding of schools was required. Codes were created randomly in Excel using the RANDBETWEEN function. Names and addresses of schools were checked for accuracy using the Schools in England database published by the Department for Education (2017) and errors were corrected to ensure that schools were assigned to the correct codes. Pupil premium figures for each school were also added onto the CSV data file using VLOOKUP codes to transfer the data from the Department for Education database. Once these changes had been made, the data was transported into R.

3.7.2.2 Processes Carried Out in R

Numbers that were stored and recognised as text were converted into numbers using the 'as.numeric' function in R. This included the pupil premium variable. Figure 2 shows the R script used to convert the pupil premium variable into numeric data.

Figure 2: R Script Showing Conversion of Pupil Premium Variable into Numeric Data

```
teachers_voice$pupil_premium <- as.numeric(teachers_voice$pupil_premium)
```

All categorical variables were stored as numeric data and were converted to categories using the 'as.factor' function in R. Figure 3 shows an example of a categorical variable being converted into a factor.

Figure 3: R Script Showing the Variable 'Subject' Being Converted Into Categories

```
teachers_voice$subject <- as.factor(teachers_voice$subject)
```

The total scores for the VoiSS, WEMWBS and GerdQ were calculated in R using the 'RowSums' function. Figure 4 shows the R script for calculating total scores on the VoiSS.

Figure 4: R Script Showing Total VoiSS Scores Being Calculated

- total_voiss <- teachers_voice[, 43:72] # Create a new dataframe from the data teachers_voice called total_voiss which includes columns 43 to 72.
- rowSums(total_voiss) #add up the total of each row in the data frame total_voiss, i.e. the VoiSS scores.
- teachers_voice$Voiss_total <- rowSums(total_voiss) # create a new variable with the total scores called Voiss_total and add to teachers_voice

As the data had been downloaded as zero-indexed, the first value on the drop-down menu that a participant selected was downloaded from BOS as the value of 0, even though the participant saw it as the correct number. Variables that did not begin at the numerical value of 0 on the drop-down menu had to be recoded. This included the variables 'hours teach a week', 'class size' and 'age'. For example, the first number that a participant could choose for 'hours teach a week' and 'class size' was 1 and the first number for age was 21. But because the numerical data had been downloaded as zero-indexed, the value 1 for 'hours teach a week' and 'class size' was coded as 0, the value of 2 coded as 1, the value of 3 coded as 2, etc. The value of 21 for 'age' was coded as 0, the age of 22 coded as 1, the age of 23 coded as 2, etc. Therefore, for the codes and values to be the same, the value of 1 had to be added to each entry for the variables 'hours teach a week' and 'class size', and a value of 21 for each value of the variable 'age'. Figure 5 shows the R script used to add the value of 21 to the variable 'age'.

Figure 5: R Script Showing the Recoding of the Variable 'Age'

teachers_voice$age <- teachers_voice$age_raw +21

Once the changes had been made to the data in R, the data was saved again as an Excel spreadsheet. Statistical analysis was carried out using this file, called 'Teachers_voiceR'.

3.7.2.4 Missing Data

When doing regression analysis in R, a complete case analysis is carried out, so any observations that have missing data are excluded from the analysis. However, leaving out participants due to missing data may lead to bias and loss of power due to reduced sample size (Bell and Fairclough, 2014). Consequently, it was necessary to consider how to handle the missing data. Data can be missing for three reasons. Missing completely at random (MCAR) is a term used to relate to missing data that is not thought to be related to any of the variables being analysed (Bhaskaran and Smeeth, 2014). This creates a loss of power but is less likely to create a bias (Kang, 2013). Missing data can be missing at random (MAR), where there is a systematic relationship between the missing values and the observed variables, in other words, the missing values can be explained by the other variables. For example, answering the WEMWBS scale may be conditional on sex and age. The Care Quality

Commission (2016) found lower response rates to a mental health survey for males and those under 35. Finally, the data that is missing is not random, called Missing not at random (MNR) occurs when there is a reason for the missing data. For example, those who take illegal drugs may not answer questions on drug use for fear of being prosecuted.

To account for the possibility of bias and loss of power, imputation was considered. Imputation involves substituting the missing values with estimates, therefore allowing all cases to be analysed (Kang, 2013) and has the potential to increase the validity of a study (Sterne et al., 2009). However, it does have some disadvantages. It assumes that the data is normally distributed, and the missing data is random. Using multiple imputation on skewed non-random data will introduce bias (Sterne et al., 2009). Additionally, it involves complicated statistical procedures, requiring specialist statistical assistance. It was felt that the time and expertise involved in the procedure were beyond the scope of the study. Therefore, this study used a complete case analysis.

3.7.3 Population Bias

A goodness of fit chi-squared analysis was carried out to determine whether there were biases in the study participants and thus whether the sample was representative of all teachers working in England. Bias was tested for age, gender, ethnicity, type of school and school region. Figures for teachers working in England were taken from the School Workforce Census 2017 (Department for Education, 2018c). However, this census only includes teachers at state-funded schools and does not include sixth form colleges. Therefore, the bias analysis did not include participants from fee-paying schools and sixth form colleges. As the bias tests involved a number of parallel tests, a Bonferroni correction was applied. A Bonferroni corrects the p-value for parallel tests to ensure that false positives are avoided (Weisstein, 2020). To do this the p-value is divided by the number of tests being carried out. As the number of tests was six, the p-value of 0.05 was divided by 6.

3.7.4 Internal Consistency

Internal consistency determines the extent to which items on a test measure the same underlying factor (Revelle, 1979). If the items do measure the same factor, then an individual would be expected to answer all the questions on the test in a similar way. The VoiSS has been tested for internal consistency in populations of voice patients and has shown good internal consistency (Webb et al., 2007). The WEMWBS has also shown high internal consistency with general populations in the UK (Fat et al., 2017; Tennant et al., 2007). However, the internal consistency of a test is only valid for the specific population it has been tested with and should, therefore be measured again when using the test with other populations (Streiner, 2003). It was thus necessary to measure the internal consistency of the VoiSS, WEMWBS and GerdQ to ensure they were valid tools for using with teachers working in England.

Internal consistency was measured using Cronbach's Alpha. Cronbach's alpha establishes whether the factors are linked by measuring the correlations between items on the test (Cronbach, 1951). Cronbach's Alpha allows the researcher to determine the reliability of tools in the study population (Cronbach, 1951). Cronbach's Alpha gives a score between -1 and 1, with positive values indicating consistent agreement and negative values consistent disagreement. Different authors have presented different alpha values as being acceptable for research and clinical purposes. Nunnally and Bernstein (1994) have suggested that 0.80 should be the minimally acceptable value for research purposes and 0.90 for clinical purposes. Bland and Altman (1997) suggest that values between 0.70 and 0.80 are acceptable for research whereas for clinical purposes alpha should be at least 0.90, with 0.95 being desirable. Therefore, in this research values of over 0.80 were deemed acceptable.

3.7.5 Univariate Linear Regression

A univariate linear regression was carried out in R to test the associations of the variables of interest and their associations. Carrying out a univariate analysis allows an exploratory investigation of the variables of interest to see how they are associated with the outcome variable without considering other variables.

3.7.6 Multiple Linear Regression

A multiple linear regression model was carried out where all the variables of interest were fitted into the same model. Multiple linear regression allows the researcher to determine the association of each variable to the outcome variable while accounting for other variables. This is an analysis that has one level of analysis and therefore does not account for the clustering that may take place amongst teachers who teach at the same schools and thus school effects. The variables of interest were all fitted together in R. Figure 6 shows the R script for the model fit.

Figure 6: R Script Showing Multiple Linear Regression

```
multiple_lin_mod <- lm(Voiss_total ~ pupil_premium + age + sex_at_birth + WarwickEd_total + gerdq_total + smoking + voice_training + subject + talk_louder + background_noise + respiratory_infect + asthma + yrs_teaching + hrs_teach_wk + class_size_cat, data = teachers_voiceR)
```

3.7.7 Multilevel analysis

A multilevel analysis was undertaken using a random-effects model (Snijders, 2005). The model takes into account school-level effects, as the clustering of teachers within schools may affect VoiSS scores outcomes. Incorrect inferences can occur if clustering is not taken into account (De Leeuw et al., 2008). A two-level model was fitted with teachers being at the first level and schools at the second level. Figure 7 shows the R script for the model fit. The school-level data is represented by the code (1|school_code).

Figure 7: R Script of Multilevel Model

```
multivariable.model2 <- lmer(Voiss_total ~ age + factor(sex_at_birth) + WarwickEd_total + pupil_premium + gerdq_total + factor(smoking) + factor(voice_training) + factor(subject) + factor(talk_louder) + factor(background_noise) + factor(respiratory_infect) + factor(asthma) + yrs_teaching + hrs_teach_wk + factor (class_size_cat) + factor(class_size_cat) + (1|school_code), data = teachers_voiceR, REML=FALSE)
```

3.7.8 Assumptions

Once the multiple linear regression and multilevel models had been fitted, they were checked to see if they met the assumptions of:

- independence of observations
- linearity
- homoscedasticity
- multicollinearity
- normality
- no outliers

For assumptions that were not met, relevant changes were made to the data. These will be discussed further in the results chapter.

3.7.8.1 Independence of Observations

Independence of observations assumes that there is no correlation among units of analysis (in this case teachers). Autocorrelation occurs when the residuals between participants are not independent from each other. For a linear regression to take place there should be little or no autocorrelation in the residuals and thus an independence of observations (Hagger-Johnson 2014). A Durbin Watson test was used to establish whether there was autocorrelation of the residuals in the data. The Durban Watson statistic establishes autocorrelation between adjacent observations (Laerd Statistics, 2015). A Durbin Watson score of less than 2 indicates positive autocorrelation and values between 2 and 4 indicate negative autocorrelation. Values of under 1 and above 3 are thought to be problematic (Field, 2013). As teachers from the same school were most likely to be similar to one another, the data was arranged by school for the analysis.

3.7.8.2 Linearity

The assumption of linearity requires that there is a linear relationship between the outcome variable and the predictor variables. This was established by plotting residuals of the multiple regression model against predicted values. In order for the assumption to be met, the line through the plot should be horizontal or approximately horizontal. If the assumption of linearity is violated then a multiple linear regression is not the appropriate type of model for the data, and non-linear models need to be considered (Kelly and Bolin 2013). Linearity was established by visually assessing the residuals versus fitted values plot of the multiple linear regression.

3.7.8.3 Homoscedasticity

Homoscedasticity assumes that the variance of the observed values of the dependent variable is constant. If this assumption is violated it can indicate the standard errors are biased and the coefficients are less precise (Yang et al., 2019). Homoscedasticity was measured by plotting the residuals of the dependent variable against fitted values and the square root of the standardised residuals against the fitted values. In order for homoscedasticity to be present, the values must be equally spread along the plotted line with no pattern in the residuals (Kelly and Bolin 2013). Homoscedasticity was established by visually assessing the plots and carrying out a Breusch-Pagan test. A Breusch-Pagan test is a statistical procedure that measures the variance of the errors in a model. If the p value of the test is less than 0.05 it indicates that heteroscedasticity is present.

3.7.8.4 Multicollinearity

Multicollinearity indicates a linear relationship between independent variables (Silvey, 1969). Multicollinearity can lead to large standard errors indicating large sampling variability which in turn leads to less precise coefficients (Mason, 1987; Mela and Kopalle, 2002). In order to test for multicollinearity a Pearson's correlation was carried out to look at the correlation of the independent variables, as a high correlation can imply multicollinearity (Alin, 2010). In order to dismiss multicollinearity, it is recommended that correlations are no larger than 0.8 (Berry et al., 1985). However there can still be variables that are not well correlated that suffer from multicollinearity, so it is necessary to use other methods to diagnose (Alin, 2010). In order to do so a variance inflation factor (VIF) analysis was carried out. This can test whether there is a linear relationship between two independent variables. Different thresholds for VIF have been recommended in the literature. Hair et al. (2014) suggest a threshold of 10 while other researchers recommend a threshold as low as 3.3 (Kock et al., 2012).

3.7.8.5 Outliers and influential data points.

An outlier is an observation that differs significantly from other data points and has a large residual (Ramaswamy et al., 2000). It can indicate that there is an error in the data or a variability in the population which may affect the model. Any case that has a much larger influence than the others should be investigated for errors (Kirkwood and Sterne, 2003), although it must be noted that outliers are generally not a problem for data sets with large sample sizes (Bruce et al., 2020). Outliers were detected by visually inspecting the residual plots of the multivariate models and looking at cases that deviated greatly from the others. Influential data points were assessed by calculating Cook's distance. Cook's distance helps identify influential cases in a regression model and is calculated by removing each data point individually from the model and refitting the regression model without it (Cook, 1977). They are not necessarily values with large residuals but

those that but are likely to change the regression (Bruce et al., 2020). Cook's distance was evaluated by visually inspecting a plot of standardised residuals against leverage points.

3.7.8.6 Normality

Normality assumes the residuals are normally distributed (Bruce et al., 2020). This means that there is little or no skewness in the data and that there is little or no kurtosis. In order to test this assumption visual inspection of a Quantile Quantile (Q-Q) plot and histogram was carried out. A QQ plot plots the quantiles of the observed values of the residuals against the quantiles from a normal distribution. In order for normality to be assumed there should be no skewness in the histogram and the QQ plot should plot a straight diagonal line or almost a straight line. If normality is violated then it is widely considered that the data should be transformed, for example by using the log or square root. However some authors argue that if sample sizes are sufficient, Schmidt and Finan (2018) suggests 10 observations per variable, then violating the assumption of normality will not impact bias and will give valid results (Schmidt and Finan, 2018; Yang et al., 2019). Indeed Schmidt and Finan (2018) suggest that transforming the data to achieve normality, can cause bias when using measurement scales designed for clinical use. Furthermore, there may be difficulties in interpreting data by clinicians that has been transformed. Therefore, the data will not be transformed even if the normality assumption is violated.

3.7.9 Interaction Terms

An interaction occurs when the effect of one independent variable on the outcome variable depends on another independent variable (Williams, 2015). Interactions need to be considered , otherwise the results of associations may be incorrect (Jewell, 2004). Observing interactions can also be beneficial as they can help target clinical interventions (Jewell, 2004). Interaction terms were added to the multilevel model. As this study is primarily interested in wellbeing, the interaction between wellbeing and other independent variables were explored. Two-way interactions were assessed as higher order interactions, i.e. those between more than two independent variables, are complicated to interpret (Kontopantelis et al. 2018). An interaction term was added to the multilevel model and a nested model analysis using anova compared the multilevel model with and without the interaction. A p value of over 0.05 indicates there is no evidence that the interaction makes a difference to the model.

3.8 Conclusion

The research that was carried out was aligned with a pragmatist's philosophy, in which the research problem was considered and evaluated, and the methodology that was most appropriate for the research problem was designed. Therefore, as this research is interested in looking at whether

wellbeing is associated with voice symptoms, as well as seeing whether there are other risk factors associated, a quantitative approach was conducted using a cross sectional design. All schools in England were invited to take part in an online web-based survey, and therefore it is hoped that results from this study will be generalisable to all teachers in England. The statistical approach for the study was a multiple linear regression using a multi-level design to take into possible school effects. The results of the statistical analysis will be discussed in the following chapter.

Chapter 4: Results

4.0 Introduction

This chapter outlines the results of the analysis undertaken. The analysis was carried out in order to determine:

1. Whether mental wellbeing, as measured by scores on the Warwick Edinburgh Mental Wellbeing Scale (WEMWBS) (Tennant et al., 2007) was associated with voice symptoms as measured by scores on the Voice Symptom Scale (VoiSS) (Deary et al., 2003).
2. Which of the remaining variables of interest were associated with scores on the VoiSS. These were age, sex, gastroesophageal disease, number of years teaching, number of hours teaching per week, talking against background noise in the classroom, speaking louder than normal, class size, subject taught, voice training, diagnosis of asthma, respiratory infection in the last 30 days and smoking status.

The analysis is divided into four sections. The first section details the baseline characteristics of the variables of interest in the teaching sample that participated and gives an overview of the descriptive statistics. It discusses missing data and describes how this was approached. Section 2 presents the biases in the study population. The age, gender, ethnicity, type of school and school region in the study population were compared to the proportion of teachers working in England at the time of the study was carried out (November 2017 – February 2018), using the Department of Education's 2017 School Workforce Census (Department for Education, 2018b). A chi-squared goodness of fit analysis was used to calculate the bias of the study sample. Section 3 shows the internal consistencies of published questionnaires used within the survey; the Voice Symptom Scale (VoiSS) (Deary et al., 2003), Warwick Edinburgh Mental Well Being Scale (WEMWBS) (Tennant et al., 2007a) and the GerdQ questionnaire (GerdQ) (Jones et al., 2009) using Cronbach's Alpha.

The final section presents results from linear regressions to determine associations between the VoiSS, the WEMWBS and other variables of interest. Initially a univariable analysis was carried out for each variable of interest to determine which variables were significantly associated at the 0.05 level with the VoiSS, without accounting for any other variables. A multiple linear model was then carried out where all the variables of interest were put in the model. In order to account for possible school effects a multilevel analysis was then carried out. Finally, interactions between wellbeing and other variables were explored to see if there were any statistically significant interactions that affected the results of the associations.

4.1 Participant Characteristics

Tables 7 to 12 show the participants' characteristics. Table 7 includes the demographics of the participants, table 8 shows their teaching characteristics, table 9 shows the VoiSS scores for the participants and their vocal training, table 10 includes the mental well-being of the participants as measured by the WEMWBS, table 11 looks at the health factors and table 12 the vocal behaviours of the participants. For continuous variables with a normal distribution the mean and standard deviation are reported and the median and interquartile ranges are reported for those without a normal distribution. A distribution is considered to be normal when the mean and median are similar to each other and located at the centre of a distribution. Categorical variables are presented with the number of responses for each category and their percentages.

4.1.1 Demographics

Table 7: Participant Characteristics: Demographics.

All characteristics are reported as numbers (percentages) unless otherwise stated.

Characteristics	**All participants (n=1205)**
Age (years), mean (sd)	38.98 (10.65)
Sex	
Female	967 (80.25)
Male	233 (19.34)
Prefer not to say	1 (0.08)
Missing data	4 (0.33)
Ethnicity	
Black African	1 (0.08)
Black Caribbean	10 (0.83)
Chinese	3 (0.25)
Indian, Pakistani, Bangladeshi	22 (1.83)
Mixed White/Asian	7 (0.58)
Mixed White/Black	8 (0.66)
Other Asian	2 (0.17)
Other Ethnicity	5 (0.41)
Other mixed	6 (0.50)
Other White	38 (3.15)
Prefer not to say	18 (1.49)
White British/Irish	1082 (89.79)
Missing data	3 (0.25)

4.1.3 Teaching Characteristics

Table 8: Teaching Characteristics

Deprivation of school measured by Pupil Premium, median (interquartile range)	21 (11.5 – 35.4)
Years teaching median (interquartile range)	10 (5-19)
Average number of hours teach per week mean(sd)	21.19 (8.55)
School level	
Primary	627 (52.03)
Secondary	578 (47.97)
Type of School (state funded/independent)	
State funded	1141 (94.69)
Independent Fee paying	64 (5.31)
Type of school (mainstream/special/PRU)	
Mainstream	1157 (96.02)
Special school	39 (3.24)
Alternative provision/PRU	9 (0.75)
Subject Taught	
Primary Classroom	567 (47.05)

Arts and Humanities	255 (21.16)
Science and mathematics	200 (16.6)
Performing Arts	76 (6.31)
Social Science	44 (3.65)
Physical Education	27 (2.24)
Other	31 (2.57)
Missing data	5 (0.41)
Class size	
1-28	697 (57.84)
29 and over	500 (41.49)
Missing data	8 (0.66)

4.1.4 Voice

Table 9: VoiSS Scores

VoiSS (Total score) median (interquartile range)	20 (11-29)
Voiss (Total score) mean (sd)	22.54 (15.47)
VoiSS (Impairment scale) median (interquartile range)	11 (6-17)
VoiSS (Emotional scale) median (interquartile range)	0 (0-2)
VoiSS (Physical scale) mean (sd)	8.19 (4.68)
Diagnosed with a Voice Problem	
Yes	117 (9.71)
Under investigation	23 (1.91)
No	1065 (88.38)
Voice Training	
No	825 (68.46)
Yes – professional development	62 (5.15)
Yes – teacher training	276 (22.90)
Yes – teacher training and professional development	41 (3.40)

4.1.5 Well Being

Table 10: Scores on the WEMWBS

Warwick Edinburgh Well Being Scale (total scores) mean (sd)	42.90 (7.85)

The scores on the Warwick Edinburgh Mental Well Being Scale ranged from a low of 14, which was the lowest possible score, to 65.64. The wellbeing for this study's teaching population was lower than the general UK population. Tennant et al. (2007) reported a median of 51 (95% CI 51-52) with a general population in Scotland, and the Health Survey for England 2017 reported an average score of 49.85 (Fuller et al., 2018).

4.1.6 Health variables

Table 11: Health variables

Likelihood of having Gastroesophogeal reflux disease as measured by GerdQ questionnaire (total scores) mean (sd)	6.64 (1.99)
Smoking status	
Every day	36 (2.99)
Never smoked	858 (72.22)
Occasionally	35 (2.90)
Used to smoke	261 (21.66)
Missing data	1 (0.08)
Respiratory Tract infection in last 30 days	

Don't think so	460 (38.17)
Yes confirmed by Dr	157 (13.03)
Yes not confirmed by Dr	584 (48.46)
Missing data	4 (0.33)
Diagnosis of asthma	
Yes	319 (26.47)
No	882 (73.20)
Missing data	4 (0.34)
Missing data	1 (0.08)

Table 12: Vocal Behaviours

Last 30 days had to speak louder than normal in the classroom	
Never	76 (6.31)
Sometimes	637 (52.86)
Most of the time	364 (30.21)
Always	126 (10.46)
Missing data	2 (0.17)
Last 30 days spoke over background noise in the classroom	
Never	461 (38.26)
Sometimes	490 (40.66)
Most of the time	170 (14.11)
Always	79 (6.56)
Missing data	5 (0.41)

4.2 Missing data

There were 207 missing responses in total in the dataset. Table 13 gives a breakdown of the variables and total number of missing values for each variable of interest. The variables with the largest amount of non-responses included years teaching (90 missing values), Warwick Edinburgh (22 missing values), Hours teaching (17 missing values) and pupil premium (17 missing values). Only the predictors had missing responses, as participants had to fill in all of the answers to the VoiSS in order to proceed to the rest of the questionnaire.

Table 13: Non-Responses for Variables

Variable	**Number of NAs**
Smoking	1
Voice Training	1
Talk louder	2
Sex	4
Respiratory infection	4
Asthma	4
Subject taught	5
Background noise	5
Class size	8
Age	11
GerdQ total	16
Pupil Premium	17
Hours teach a week	17
Total Warwick Edinburgh scores	22
Years teaching	90
Total non-responses	207

4.3 Population Bias

This section will show the biases in the study population using a goodness of fit chi squared analysis. When a Bonferroni correction was applied to the level of significance (initially at alpha = 0.05), the resulting alpha was 0.0083. Therefore, only p values below this value were considered to be statistically significant.

4.3.1 Age

Table 14 shows the number and proportions of teachers of each age group in state funded schools as collected by the government schools census for 2017 as well as the observed numbers of participants in the study population and the expected number based on the government census proportions. There were 10 non-responses for age, so this left 1122 participants to analyse. The figures in the school workforce census for age categories include state funded nursery school teachers, which our study sample does not include. These numbers, although they are not expected to greatly alter figures may affect totals and proportions. Therefore, this needs to be considered when interpreting results.

Table 14: Total and Proportion of Teachers in Age Categories

Figures from School Workforce census 2017 (numbers in thousands).

Age	Total number of teachers in workforce (numbers in thousands)	Proportion in each category	Expected numbers in dataset	Observed numbers in data set	Observed proportions
Under 25	28.6	0.06	72	81	0.07
25-29	81.4	0.18	204	201	0.18
30-34	78.1	0.17	195	168	0.15
35-39	69	0.15	173	160	0.14
40-44	59.9	0.13	150	143	0.13
45-49	53.9	0.12	135	162	0.14
50-54	42.7	0.10	107	107	0.10
55-59	25.1	0.06	63	76	0.07
60 and over	9.3	0.02	23	24	0.02
Total	**448**		**1122**	**1122**	

The results for chi squared are:

$$\text{X-squared} = 14.66$$
$$\text{df} = 8$$
$$\text{p value} = 0.066$$

The p value is above the 0.0083 level, indicating that there is no statistically significant difference in association between the observed and expected values. This indicates there is no bias in terms of age in the study sample.

4.3.2 Gender/Sex

Table 15 shows the total number and proportion of male and female teachers working in state funded schools in England and the observed and expected number in the sample. There was 1 participant who answered, 'prefer not to say' and 4 nonresponses. This left 1,127 participants to analyse. The figures from the School Work Census record gender whereas the data in this study

records sex at birth. Although this is not expected to significantly affect the figures this needs to be considered when interpreting the results.

Table 15: Total Number and Proportions of Male and Female Teachers in England

Figures taken from the School workforce census 2017 (numbers in thousands). Observed and expected numbers of participants of each category in dataset.

Gender	Total numbers of teachers	Proportion	Expected numbers in dataset.	Observed numbers in dataset	Observed proportions
Male	120.7	0.25	276	211	0.19
Female	371.2	0.75	851	911	0.81

The results for chi squared are:
$\text{X-squared} = 19.67$
$\text{df} = 1$
$\text{p value} = <0.001$

The p value demonstrates that there is a statistically significant difference in the observed and expected values for male and female teachers. This demonstrates that the study sample of participants from state funded schools is biased towards females. Although the figures indicate that the men in the dataset are unrepresented, they make up 19% of our sample rather than the expected 25% so this should not be problematic. There should be enough in our sample to be able to generalise to other male teachers.

4.3.3 Ethnicity

Table 16 shows the percentages, proportions, observed, and expected values for the ethnicity of teachers in England. Expected numbers are rounded to the nearest whole number. The figures from the school census for ethnicity include state funded centrally employed teachers. Centrally employed teachers include supply teachers and peripatetic teachers who were not included in our study sample. However centrally employed teachers only account for a small proportion of the overall teaching population, about 0.84%, so including their figures is not expected to particularly affect results but needs to be taken into consideration when interpreting them.

Table 16: Totals and Proportions of Ethnic Groups among Teachers in State Funded Schools.

Ethnicity	Percentage of teachers in each category	Proportion	Expected number of participants in dataset.	Observed number of participants in dataset	Observed proportions
White British/Irish	87.3	0.873	988	1022	0.903
Any other white background	3.8	0.038	43	35	0.031
Mixed White and Black	0.4	0.004	5	7	0.006
White and Asian	0.3	0.003	3	5	0.004
Any other mixed background	0.5	0.005	6	6	0.005
Indian/Pakistani/Bangladeshi	3.5	0.035	40	20	0.018
Chinese	0.2	0.002	2	3	0.003
Any other Asian Background	0.6	0.006	7	2	0.002
Black Caribbean	1.0	0.01	11	8	0.007
Black African	0.8	0.008	9	1	0.001
Any other ethnic group	0.9	0.009	10	5	0.004
Refused	0.7	0.007	8	18	0.016

Total			1132	1132	

As the proportions and number of participants for some of the ethnic groups were so small, to aid analysis, ethnicity was divided into white and other ethnic groups. The proportions of the school workforce expected number of participants and observed number and proportions of participants is shown in table 17.

Table 17: Totals and Proportions of White and Other Ethnic Groups of Teachers

Ethnicity	Percentage of teachers in each category	Proportion	Expected number of participants in dataset.	Observed number of participants in dataset	Observed proportions
White	91.1	0.911	1031	1057	0.934
Other ethnic groups	8.2	0.082	93	57	0.05
Refused	0.7	0.007	8	18	0.016
Total			1132	1132	

The results for chi squared are:

$$\text{X-squared} = 27.28$$
$$\text{df} = 2$$
$$\text{p value} = <0.001$$

This value is lower than the corrected p value of 0.0083, and therefore the difference between the observed and expected values is statistically significant. When the figures are compared between observed and expected values it can be seen that White participants account for 93% of the sample compared with 91% expected. This indicates a bias towards White participants. However, when looking at table 16 there are some ethnic groups in this study sample that have a similar or a greater proportion and numbers than those in the School Workforce Census, thus indicating that the results can be generalisable to these groups. These include Mixed White and Black, White and Asian, Any other mixed background Black Caribbean and Chinese. Other ethnic groups however had smaller proportions than the national census so results may need to be applied cautiously to these groups. These include Indian/Pakistani/Bangladeshi, Any other Asian background, Black African and Any other ethnic group.

4.3.4 Type of school

Table 18 shows the numbers and proportions of teachers working in England in different types of school, including primary, secondary, special schools and alternative provision/PRU. The numbers of teachers in the School workforce census working at the Primary level include nursery school teachers. This needs to be accounted for when interpreting results.

Table 18: The Number and Proportion of Teachers Working in Different Schools

School level	Number	Proportion	Expected	Observed	Observed proportions
Primary	221.1	0.49	553	586	0.52
Secondary	204.2	0.45	512	500	0.44
Special	22.8	0.05	58	37	0.03
Centrally Employed (includes PRU alternative provision)	3.8	0.01	9	9	0.01
Total			**1132**	**1132**	

The results for chi squared are:

$$\text{X-squared} = 9.61$$
$$\text{df} = 3$$
$$\text{p value} = 0.022$$

The p value is above the 0.0083 level and therefore indicates that there are no statistically significant differences between observed and expected values and therefore no bias in this study's sample of teachers.

4.3.5 School Region

Table 19 shows the total and proportion of teachers working in each region of England according to the School Workforce Census (2017). It also shows the observed numbers in the dataset compared to the expected number. 17 non-responses had not given school details, leaving 1117 participants to analyse.

Table 19: Total and Proportion of Teachers by Region of England

Observed and expected number of teachers of each category in the data set.

Region	Total	Proportion	Expected	Observed	Observed proportions
East Midlands	41424	0.083	93	89	0.080
East of England	56686	0.114	127	116	0.104
London	80383	0.162	181	42	0.038
North East	23366	0.047	52	15	0.013
North West	65858	0.132	148	75	0.067
South East	78934	0.159	178	119	0.107
South West	47412	0.095	106	99	0.089
West Midlands	55052	0.111	124	539	0.483
Yorkshire and Humber	48066	0.097	108	23	0.021
Total			**1117**	**1117**	

The Chi squared results are:

$$\text{X-squared} = 1646.4$$
$$\text{df} = 8$$
$$\text{p value} = <0.0005$$

The low p value of <0.0005 shows a statistically significant association between observed and expected values. The only region with more observed than expected values was the West Midlands, with 540 observed teachers compared to 125 expected numbers. This shows a large bias towards teachers from the West Midlands taking part in the survey. However, the observed and expected

proportions for some of the other regions are similar to each other and therefore the results can also be applied to these regions. These include the East Midlands where the observed proportion was 8% compared to an expected 8.3%, the East of England where the observed proportion is 10 % compared to 11%, and the South West with observed values of 8.9% compared to an expected proportion of 9.5%. However, the observed proportions for London, the North East, the North West, the South East and Yorkshire and Humber are much lower than the expected values and therefore this may present a problem in generalising the results to these regions.

4.3.6 Fee paying and State schools

Table 20 shows the numbers of teachers and proportions in state funded and fee-paying schools. It also shows the observed values of participants from these schools and the expected number of participants if there was no bias. The numbers of teachers from the fee-paying schools in the table is an approximation. The figures come from the Independent Schools Council census (ISCC) (Stevens et al., 2018) which does not represent every independent school in the UK. Additionally, the ISCC figures incorporate the whole of the UK rather than just England. However, the vast majority of schools in the ISCC are in England, so it is expected that this will not significantly affect the figures.

Table 20: Number and Proportion of Teachers from State Funded and Fee-Paying Schools

Type of School	Number of Teachers	Proportion	Expected	Observed	Observed proportions
State Funded	447,700	0.871	1041.7	1130	0.945
Fee Paying	66,038	0.129	154.3	66	0.055
Total			1196	1196	

The results for chi squared are:

$$\text{X-squared} = 58.0$$
$$\text{df} = 1$$
$$\text{p value} = <0.001$$

The p value shows that there is a highly significant difference between observed and expected values of participants from state and fee-paying schools. There was a bias towards teachers working at state schools completing the survey. Additionally, the proportions of observed and expected values are very different which suggests that the low number of teachers from independent schools completing the survey may be a problem and therefore results from this study should be applied to this group with caution.

4.3.7 Summary

The results from the chi square tests show that the results of this study are most valid for state funded teachers of all ages and in all types of schools, who teach in the West Midlands, are female and identify as White. Although the significance tests indicate bias, due to the similarity in proportions between the observed and expected results it should not be a problem to generalise the

results to male teachers, those of Mixed White and Black, White and Asian, Any other mixed background, Black Caribbean and Chinese ethnicity and those teaching in the East Midlands, East of England and the South West regions. However, results should be applied with caution to those teaching in independent schools, to Black African teachers, Indian/Pakistani/Bangladeshi, any other Asian background and any other ethnic group and those teaching in London, the South East, the North West, the North East and Yorkshire and the Humber.

4.4 Internal Consistency

4.4.1 Voiss

Cronbach's Alpha was 0.94 with a 95% CI of 0.94-0.95 for the VoiSS as calculated from this study population.

This indicates a strong internal consistency for the VoiSS in this study population and is therefore a suitable measure to use in the study.

4.4.2 WEMWBS

Cronbach's Alpha was 0.94 with a 95% CI of 0.94-0.95 with this study population, indicating strong internal consistency. It is therefore a good measure to use in this study.

4.4.3 GerdQ

Cronbach's Alpha was 0.19 with a 95% CI of 0.12, 0.26 indicating a poor internal consistency and therefore may not be a suitable measure to use in our study population. Figure 8 shows the printout from R showing the reliability when an item is dropped from the GerdQ. GerdQ3 and GerdQ4 are positively correlated whereas the other questions are negatively correlated with the total scale. This is because Q1, Q2, Q5 and Q6 are scored with the response '0 day' corresponding to a score of 0, '1 day' a score of 1, '2-3 days' a score of 2, and '4-7 days' a score of 3. Questions Q3 and Q4 have reversed scoring so that a response of '0 day' has a score of 3, etc. In order to improve internal consistency, the scores for each of the questions need to be in the same direction, therefore those for GerdQ3 and GerdQ4 were reversed.

Figure 8: Showing Reliability of the Scale when an Item is Dropped

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha se	var.r	med.r
GerdQ1	-0.2357	-0.1939	0.29	-0.0336	-0.1624	0.059	0.22	-0.28
GerdQ2	-0.0092	0.0059	0.40	0.0012	0.0059	0.048	0.24	-0.27
GerdQ3	0.5377	0.5519	0.67	0.1976	1.2316	0.019	0.21	0.49
GerdQ4	0.4579	0.4566	0.64	0.1439	0.8402	0.020	0.27	0.49
GerdQ5	-0.1370	-0.1617	0.31	-0.0286	-0.1392	0.053	0.23	-0.27
GerdQ6	-0.2489	-0.2487	0.28	-0.0415	-0.1992	0.058	0.24	-0.31

When questions 3 and 4 were reversed this produced a much stronger internal consistency. The Cronbach's Alpha was 0.83 with a 95% confidence interval of 0.82, 0.85. Figure 9 shows the reliability when an item is dropped from the questionnaire with the questions reversed.

Figure 9: Showing Reliability of the Scale when an Item is Dropped (Q3 & Q4 reversed)

```
        raw_alpha std.alpha G6(smc) average_r S/N alpha se  var.r med.r
GerdQ1       0.80      0.80    0.78      0.44 3.9   0.0094 0.0095  0.47
GerdQ2       0.80      0.80    0.79      0.45 4.1   0.0092 0.0165  0.47
GerdQ3-      0.81      0.81    0.79      0.46 4.3   0.0088 0.0170  0.49
GerdQ4-      0.84      0.84    0.81      0.51 5.3   0.0073 0.0039  0.49
GerdQ5       0.80      0.80    0.78      0.44 4.0   0.0092 0.0106  0.47
GerdQ6       0.81      0.81    0.79      0.46 4.3   0.0088 0.0080  0.48
```

Consequently, in the following univariable analysis the results of the total scores of the GerdQ as it was originally designed to be used and with questions 3 and 4 flipped were compared to consider which should be included in the multivariate analysis.

4.5 Simple linear regression

4.5.1 Continuous Variables

Table 21 shows the results of univariate analysis looking at the association of the VoiSS with the continuous variables of interest. Figures 10 to 16 show the plots of the associations.

Table 21: Univariate Analysis for Continuous Variables

Characteristic	Coefficient (95% Confidence interval)	P value	Standard Error
Years teaching	0.02 (-0.08, 0.12)	0.687	0.05
Hours teach per week	0.09 (-0.01, 0.2)	0.076	0.05
Warwick Edinburgh	-0.49 (-0.60, -0.38)	**< 0.001**	0.06
Age	-0.00004 (-0.083, 0.083)	0.976	0.04
GerdQ	1.61 (1.18, 2.01)	**<0.001**	0.2
GerdQ Q3 and Q4 flipped	1.41 (1.15, 1.66)	**<0.001**	0.1
Pupil Premium	0.05 (-0.01, 0.10)	0.088	0.03

Figure 10: Plot of VoiSS and Years Teaching

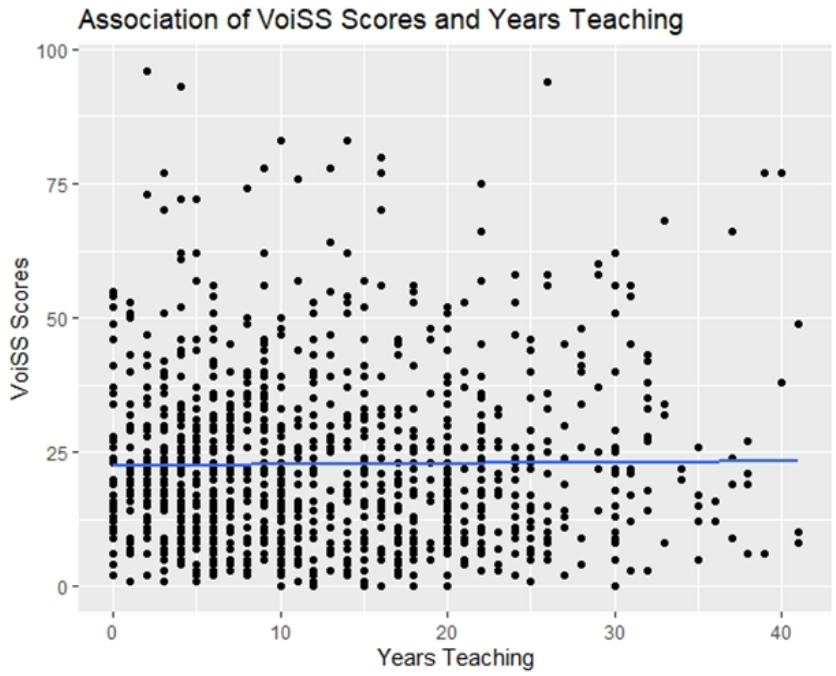


Figure 11: Plot of VoiSS and Hours Teaching per Week

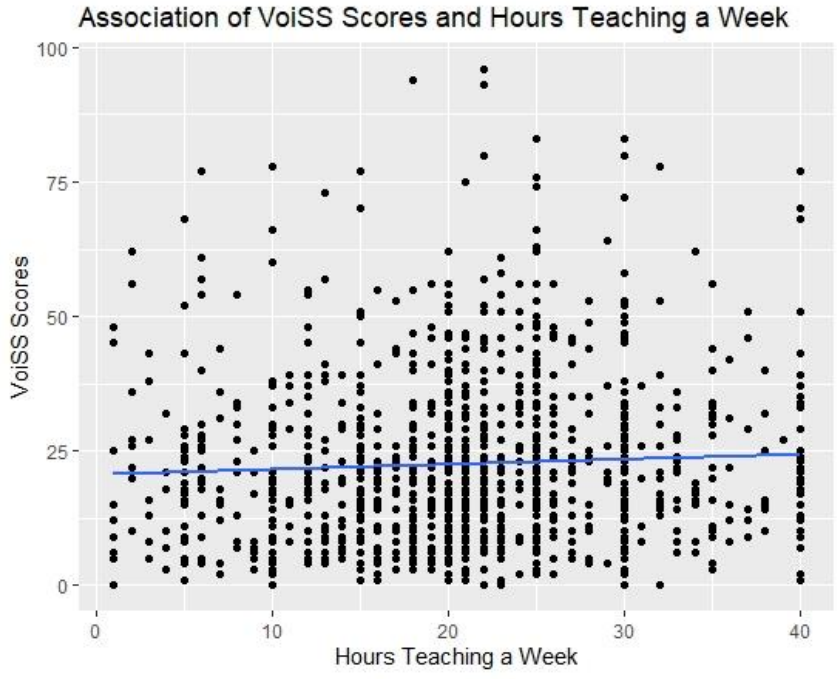


Figure 12: Plot of VoiSS and GerdQ Total

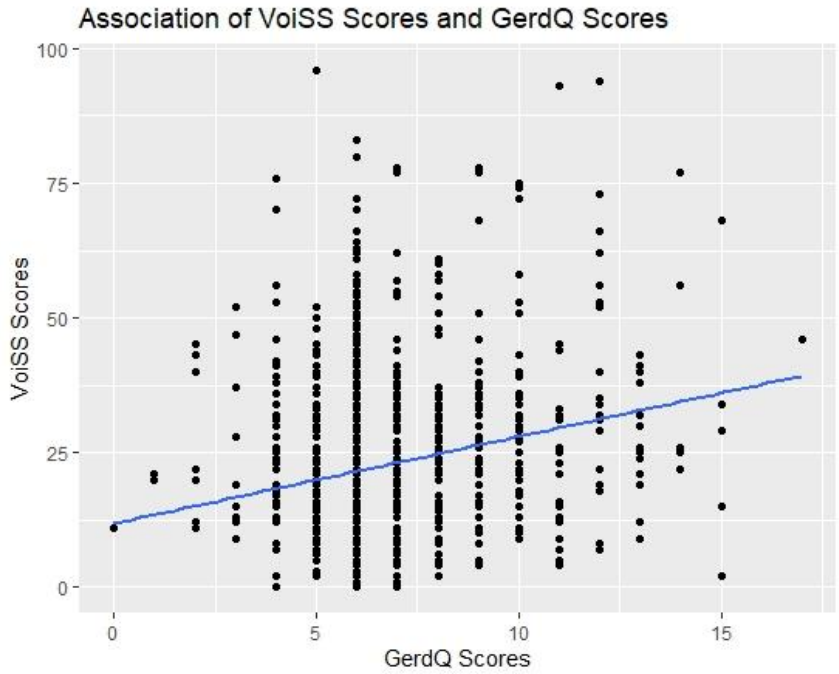


Figure 13: Plot of VoiSS and GerdQ Total with Q3 and Q4 Flipped

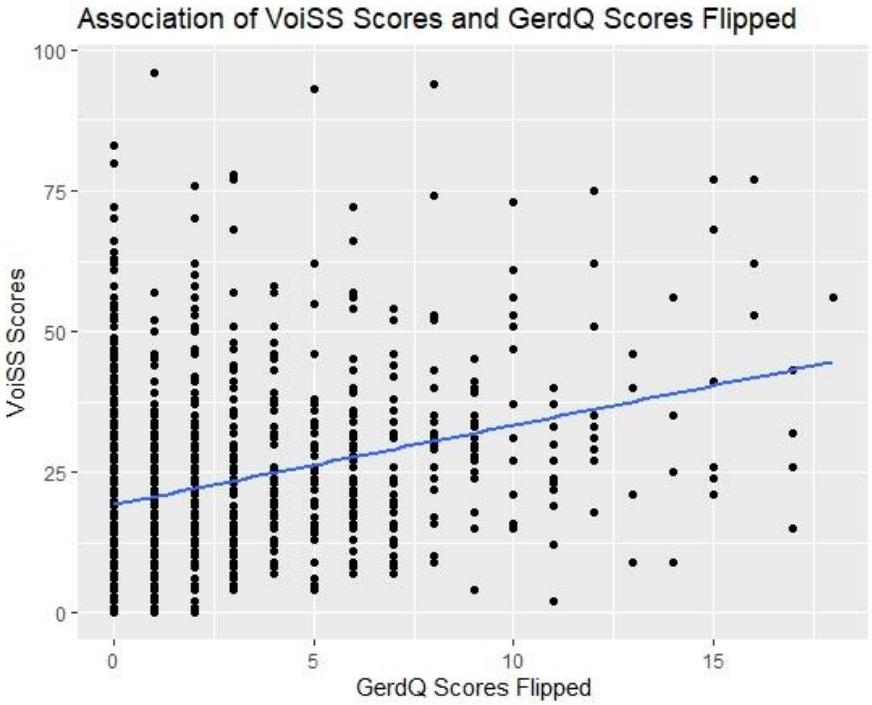


Figure 14: Plot of VoiSS and Age

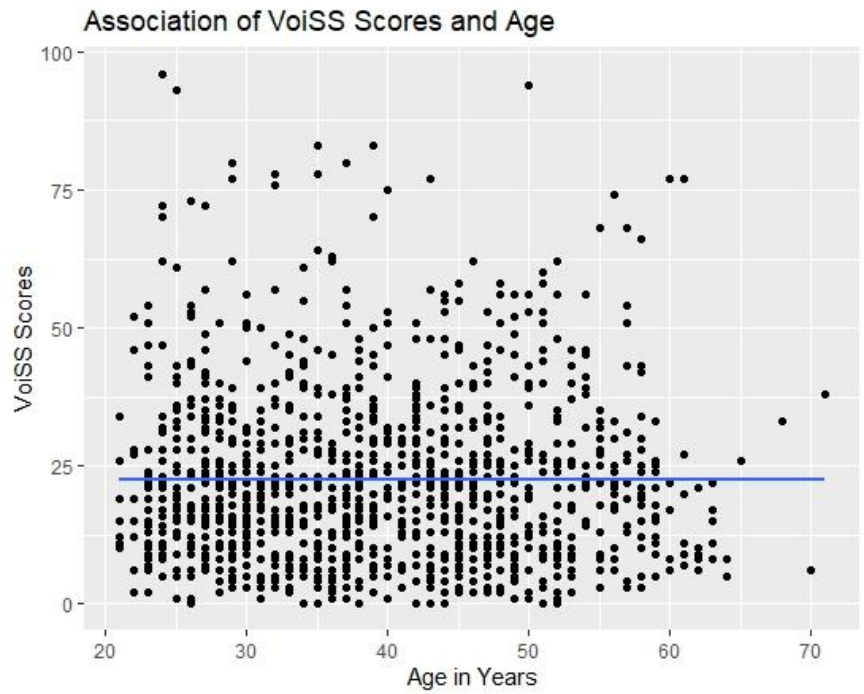


Figure 15: Plot of VoiSS and WEMWBS

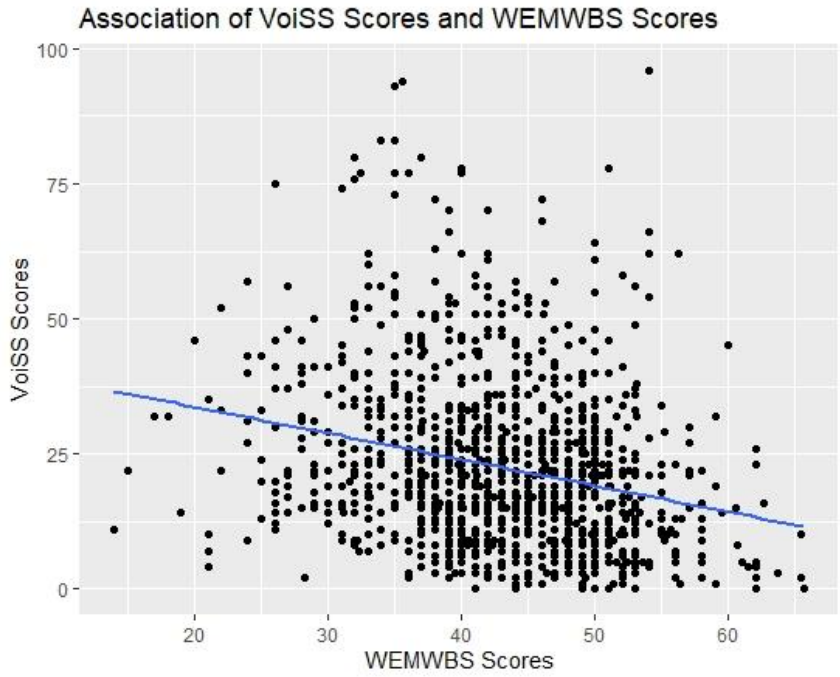
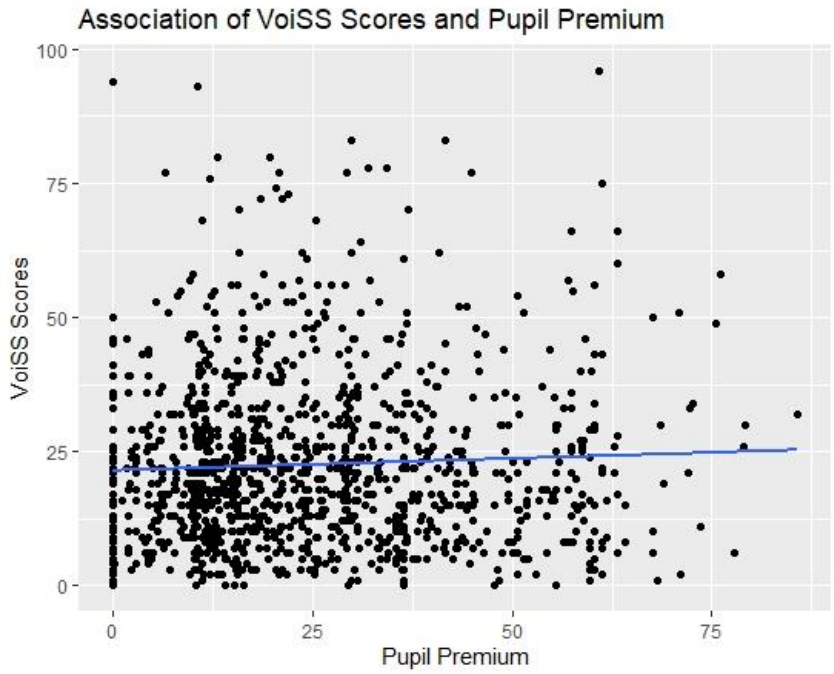


Figure 16: Plot of VoiSS and Deprivation Score (pupil premium)



4.5.2 Categorical Variables

Table 22 shows the results of the categorical variables.

Table 22: Categorical Variables

Variable	Coefficient (95% Confidence interval)	P value	Standard errors
Voice Training			
No Voice Training	Reference		
Professional Development	-1.11 (-5.14, 2.92)	0.590	
Teacher Training	-1.02 (-3.16, 1.11)	0.348	
Professional Development & Teacher Training	-3.35 (-8.21, 1.51)	0.177	
Asthma			
No	Reference		
Yes	2.71 (0.73, 4.70)	0.007	1.01
Smoking status			
Every day	Reference		
Never	-1.69 (-6.87, 3.50)	0.524	2.64
Occasionally	-0.83 (-6.40, 8.06)	0.822	3.69
Used to	-1.55 (-6.97, 3.86)	0.574	2.76
Sex			
Female	Reference		
Male	-5.18 (-7.38, -2.98)	<0.001	1.12
Prefer not to say	-20.54 (-50.69, 9.60)	0.182	15.37
Speaking over background noise			
Always	Reference		
Most of the time	-4.93 (-8.95, -0.91)	0.016	2.1
Never	-13.78 (-17.38, -10.19)	<0.001	1.8
Sometimes	-9.95 (-13.53, -6.37)	<0.001	1.8
Talk louder than normal			
Always	Reference		
Most of time	-4.84 (-7.85, -1.83)	0.002	1.5
Never	-16.00 (-20.22, -11.76)	<0.001	2.2
Sometimes	-11.79 (-14.63, -8.95)	<0.001	1.4
Respiratory infection			

No	Reference		
Yes confirmed by doctor	-5.69(-7.52, -3.87)	<0.001	0.93
Yes not confirmed by doctor	7.38 (4.74, 10.01)	<0.001	1.34
Subject			
Arts and humanities	Reference		
Primary classroom	0.37 (-1.92, 2.67)	0.75	1.2
Science and maths	-2.50 (-5.37, -0.37)	0.088	1.5
Performing arts	3.73 (-0.24, 7.71)	0.066	2.03
Social science	-3.23 (-8.19, 1.74)	0.20	0.2
PE	-1.23 (-7.38, 4.93)	0.70	3.1
Other	2.52 (-3.26, 8.31)	0.39	2.9
Class size			
29 and over	Reference		
Under 29	-4.15 (-5.92,-2.37)	<0.001	0.9

As there were no significant differences between the reference category 'no voice training' and each of the other three categories of voice training, these categories were combined into 1 category that was called 'yes'. Table 22 shows the results from the univariate analysis when there were two categories for voice training. These two categories were then used subsequently and put into multiple regression model.

Table 23: Simple Linear Regression of Voice Training

	Coefficient (95% Confidence interval)	P value	Standard Error
Voice training			
No	Reference		
Yes	-1.41 (-3.3, 0.48)	0.144	1.0

4.6 Multiple linear regression

A multiple regression was run to find associations between scores on the VoiSS and the variables of interest. Before fitting the final model testing of the assumptions of independence of observations, linearity, homoscedasticity, multicollinearity, influential data points and normality were undertaken. Changes were made to the model when appropriate.

4.6.1 Testing Assumptions.

4.6.1.1 Independence of Observations

A Durbin Watson statistic calculation was carried out to measure autocorrelation of the residuals and thus whether there was independence of observations. A Durban Watson Statistic of 2.002 indicated that the assumption of independence of residuals was met. The p value of 0.936 showing a non-statistically significant result also confirmed there was no autocorrelation. The result of the Durban Watson test is shown in figure 17.

Figure 17: Durban Watson Test

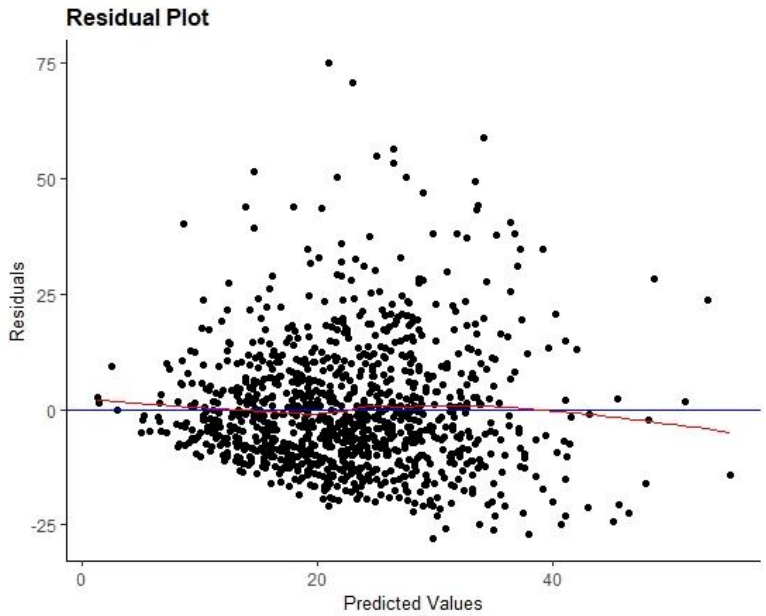
Autocorrelation D-W Statistic p-value

-0.002331654 2.001529 0.936

4.6.1.2 Linearity

Linearity was established by visually inspecting a residual versus plot. This is shown in figure 18. The red line shows an approximately horizontal line, establishing the assumption of linearity.

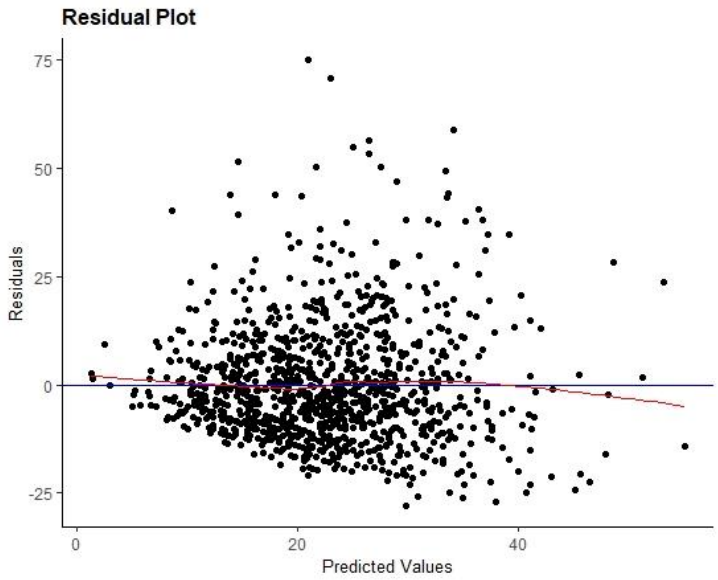
Figure 18: Plot to Show Linearity



4.6.1.3 Homoscedasticity

Figure 19 shows plots of residuals versus fitted values and the square root of standardised residuals against fitted values. In order for there to be homoscedasticity there needs to be constant variance, indicated by evenly spread points around a horizontal line. On visual inspection of the residuals vs fitted plot there was a slight funnel shape, which means that the homoscedasticity assumption may be violated. Additionally, the Scale-location plot shows a diagonal line indicating that there is not a constant variance in the data. The Breusch-Pagan test gave a p value of 0.003 indicating heteroscedasticity and confirming that the assumption of homoscedasticity had been violated.

Figure 19: Plots to Test for Homoscedasticity



4.6.1.4 Multicollinearity

In order to establish multicollinearity, the correlation between independent variables was calculated. It was established that there was a high correlation of 0.818 between years teaching and age. In order to further investigate this correlation, VIF scores were calculated, the results of which are shown in table 24. The highest VIF score was the variable 'years teaching' at 3.340, and the second highest VIF score is age at 3.296. Although many authors do not consider scores under 10 to show multicollinearity, according to Kock et al. (2012) VIFs over 3.3 may represent a problem. When looking at the VIF results along with the correlation result, it shows that years teaching and age may show multicollinearity. Additionally, years teaching has 90 missing data points and therefore using it in the analysis would reduce the sample size, creating loss of power and add possible bias. Therefore, this variable was taken out of the analysis. Following the removal of the variable 'years teaching' the assumption of homoscedasticity was retested, and a similar fit to the plots shown to figure 18 was found.

Table 24: Variables with VIF Scores

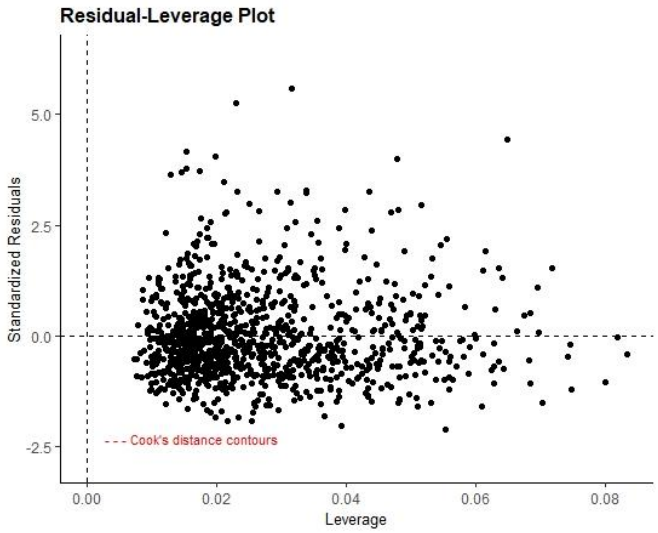
Variable	VIF score
Age	3.296
Sex at birth	1.123
Warwick Edinburgh score	1.101
GerdQ score	1.071
Smoking	1.101
Voice training	1.183
Subject taught	1.541
Talk louder	1.524

Background noise	1.443
Respiratory infection	1.107
Asthma	1.058
Years Teaching	3.340
Hours teach per week	1.150
Class size	1.173
Pupil premium	1.0602

4.6.1.5 Outliers

Outliers were identified in the residual plots. Figure 20 shows a Residuals vs Leverage plot with Cook's distance. Data points outside the red dashed line indicate those with a high Cook's distance and those that are potentially influential to the model. As can be seen in figure 20 there are no data points outside the dashed line and therefore no data points with a high Cook's distance. It was there fore not necessary to remove any data points from the model.

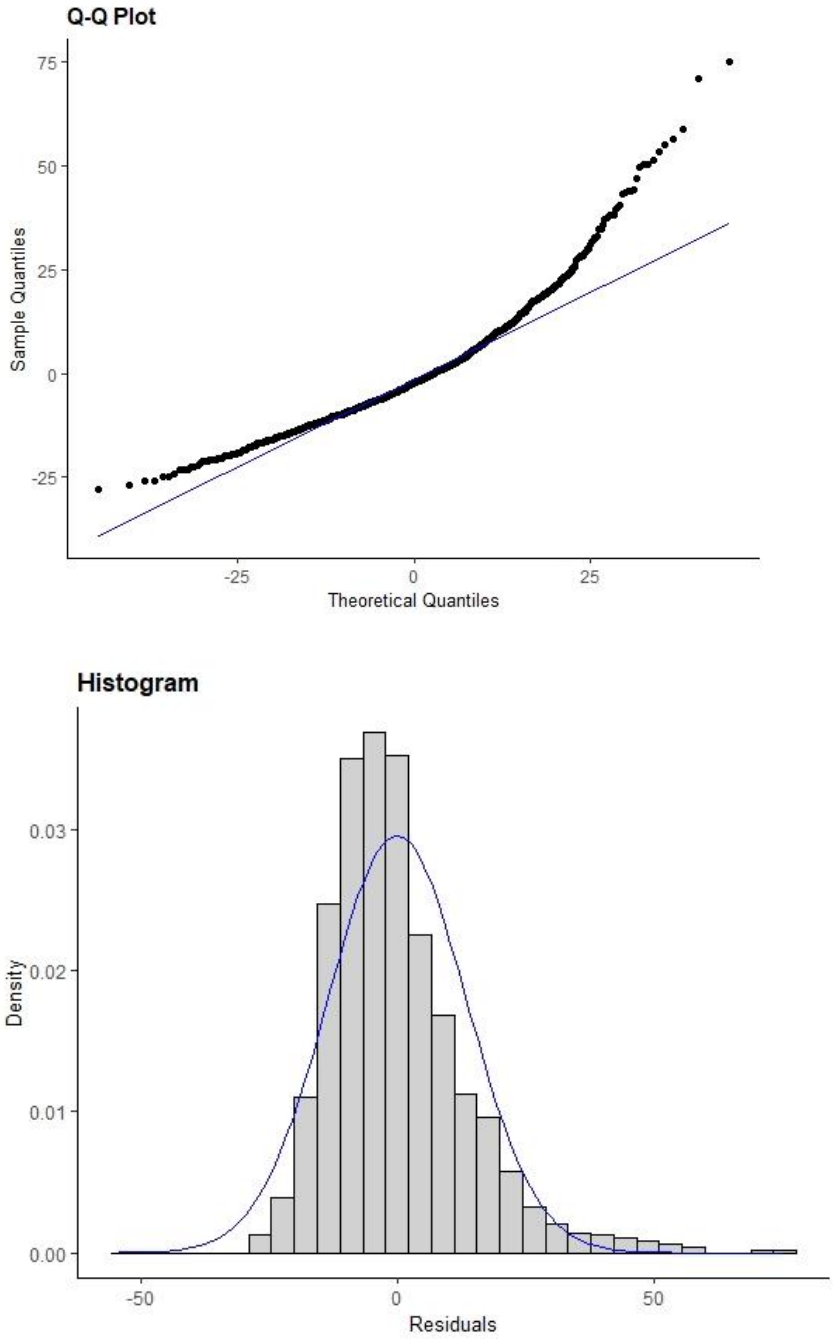
Figure 20: Plot showing Cooks Distance



4.6.1.6 Normality

To test the assumption of normality a Q-Q plot and histogram was visually analysed. Figure 21 shows the Q-Q plot and histogram. On visual inspection, it can be seen on the Q plot that the residuals curve away from the dotted line, particularly at the upper extremity and the histogram shows that the data is skewed. Thus, the assumption of normality has not been met.

Figure 21: Testing Assumption of Normality



Although the assumptions of normality and heteroscedasticity were violated, in order to aid interpretation of the results no transformation of the data took place. As discussed in the methodology chapter it is acceptable not to transform data that is not normally distributed when sample sizes are sufficient (Schmidt and Finan 2018). However, violating the assumption of homoscedasticity is more problematic as standard errors may be biased. Using robust standard errors was considered. However, using robust standard errors assumes that there is no bias in the underlying model (Freedman 2006; Zeileis 2006). In a biased model, the calculated parameters will

not be useful. Therefore, a multiple regression was carried out without adjusting for heteroscedasticity. The results of the multiple regression thus need to be interpreted with caution.

A multiple regression was run to see the association between the risk factors and scores on the VoiSS. A goodness of fit for the model using adjusted R squared was also carried out.

4.6.2 Multiple linear regression model

Table 25 shows the results of the multivariable analysis.

Table 25: Multiple Regression Analysis

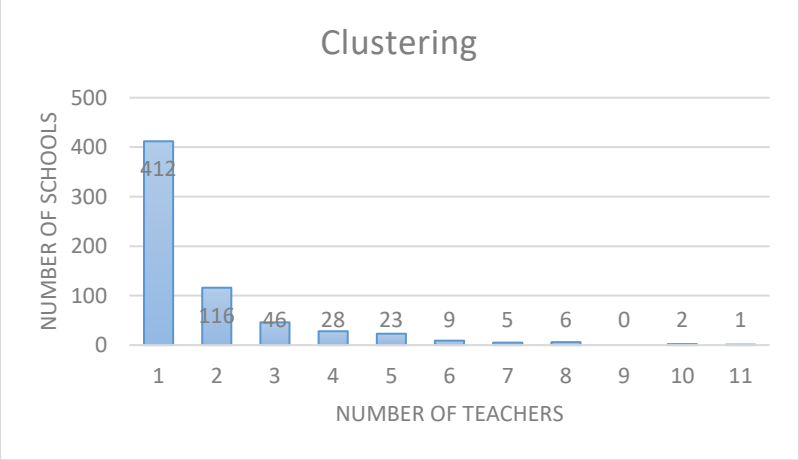
Variable	Coefficient (95% Confidence interval)	Standard Errors	P value
Warwick Edinburgh total score	-0.32 (-0.43 -0.21)	0.06	<0.001
Age	0.11(0.03, 0.19)	0.042	0.008
GerdQ total	1.28 (0.87, 1.69)	0.211	<0.001
Hours teaching per week	-0.03 (-0.13, 0.07)	0.050	0.548
Pupil premium	0.03 (-0.02, 0.08)	0.025	0.198
Sex at birth			
Female	Reference		
Male	-3.48 (-5.59, -1.37)	1.08	0.001
Prefer not to say	-22.75 (-49.45 3.96)	13.59	0.095
Smoking status			
Every day	Reference		
Never	-0.39 (-5.07 4.29)	2.38	0.872
Occasionally	-0.10 (-6.58 6.38)	3.30	0.975
Used to	-1.21 (-6.10, 3.68)	2.49	0.628
Voice training			
No	Reference		
Yes	-1.04 (-2.80 0.72)	0.90	0.248
Subject			
Arts and humanities	Reference		
Other	2.56 (-2.83, 7.96)	2.75	0.352
PE	0.52 (-5.06, 6.10)	2.84	0.856
Performing arts	-0.52 (-4.25, 3.21)	1.90	0.785
Primary classroom	-0.25 (-2.51, 2.00)	1.15	0.826
Science	-0.68 (-3.36, 2.00)	1.36	0.619
Social science	-3.56 (-8.13, 1.01)	2.33	0.321
Talk louder			
Always	Reference		
Most of the time	-2.91(-5.95, 0.13)	1.55	0.061
Never	-11.11(-15.48, -6.74)	2.23	<0.001
Sometimes	-8.14 (-11.17, -5.11)	1.55	<0.001
Background noise			
Always	Reference		
Most of the time	-0.84 (-4.80 3.12)	2.02	0.677
Never	-5.55(-9.33, -1.77)	1.93	0.004
Sometimes	-3.78 (-7.44, -0.11)	1.87	0.044
Respiratory infection			
No	Reference		
Yes confirmed by doctor	-4.03 (-5.79, -2.27)	0.90	<0.001
Yes not confirmed by doctor	6.50 (4.00, 9.01)	1.28	<0.001
Asthma			
No	Reference		
Yes	0.45 (-1.40, 2.29)	0.941	0.633
Class size			
29 and over	Reference		
Under 29	-2.17 (-3.93, -0.42)	0.895	0.015

The model significantly predicted Voiss scores ($p<0.001$). The adjusted R squared score was 0.2396 suggesting that the model accounts for 23.96% of the variance.

4.6.3 Multilevel model

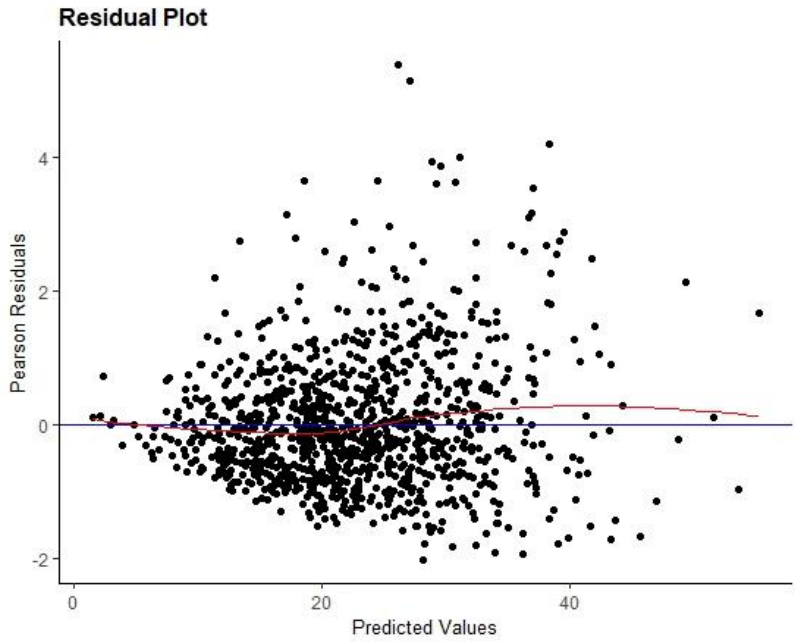
A multilevel model was carried out to take into account clustering of teachers in schools and to see if it had any effect on the coefficients. The unit of analysis was teachers and the nesting variable was schools. The school level variable used to account for clustering was 'school_code'. Figure 22 shows how many teachers there were from each school participating in the study. There were teachers from 648 schools that participated in the study. The majority of those schools, 412, only had one teacher participating. The greatest number of teachers from one school was 11 participants.

Figure 22: Number of Teachers per School



The assumptions for linearity, homoscedasticity and normality were assessed with a residual plot and a QQ plot (figure 23). Similar to the multiple linear regression model the linearity assumption was met, with an almost horizontal red line on the residual plot. The funnel shape in the residual plot suggests heteroscedasticity and the curve away from the diagonal line on the QQ plot indicates, as with the multiple linear regression, the data is not normally distributed. However, to aid the interpretation of results no transformations of the data were made. Because of the violation of homoscedasticity, the results need to be interpreted with caution.

Figure 23: Plots for Assessing Linearity, Normality and Homoscedasticity



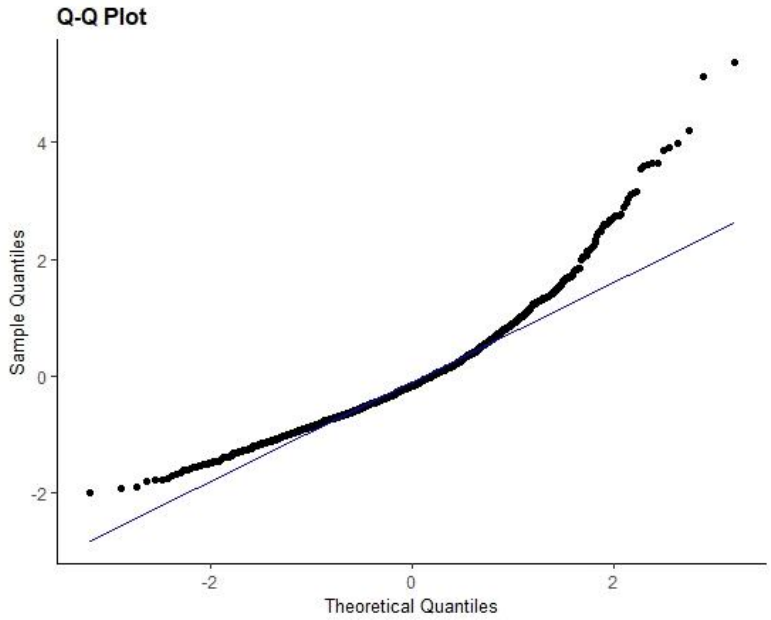


Table 26 shows the coefficients, 95% confidence intervals and p values of the variables of interest once clustering had been accounted for.

Table 26: Multilevel Model

Variable	Coefficient (95% Confidence interval)	Standard Errors	P value
Hours teach (per week)	-0.04 (-0.13, 0.06)	0.050	0.474
Warwick Edinburgh Mental Wellbeing Scores	-0.31 (-0.41, -0.20)	0.055	< 0.001
Age (in years)	0.10 (0.02, 0.18)	0.042	0.015
GerdQ Questionnaire scores	1.29 (0.87, 1.70)	0.210	<0.001
Pupil Premium	0.042 (-0.008, 0.09)	0.026	0.100
Voice Training			
No	Reference		
Yes	-0.995 (-2.76, 0.77)	0.897	0.268
Asthma			
No	Reference		
Yes	0.35 (-1.49, 2.20)	0.942	0.708
Smoking status			
Every day	Reference		
Never	-0.44 (-5.11, 4.23)	2.377	0.853
Occasionally	0.08 (-6.38, 6.54)	3.291	0.981
Used to	-1.32 (-6.19, 3.56)	2.484	0.596
Sex			
Female	Reference		
Male	-3.48 (-5.59, -1.37)	1.072	0.001
Prefer not to say	-23.79 (-50.34, 2.79)	13.531	0.079
Speaking over background noise			
Always	Reference		
Most of the time	-0.79 (-4.74, 3.17)	2.017	0.697
Never	-5.39 (-9.18, -1.61)	1.928	0.005
Sometimes	-3.46 (-7.13, 0.22)	1.872	0.065
Talk louder than normal			
Always	Reference		
Most of time	-2.85 (-5.88, 0.19)	1.548	0.066
Never	-11.35 (-15.73, -6.98)	2.228	<0.001
Sometimes	-8.23 (-11.26, -5.20)	1.546	<0.001
Respiratory infection			
No	Reference		
Yes, confirmed by a Doctor	-4.15 (-5.91, -2.39)	0.895	<0.001
Yes, not confirmed by a Doctor	6.26 (3.76, 8.77)	1.276	<0.001
Subject			
Arts and Humanities	Reference		
Primary classroom	-0.35 (-2.66, 1.96)	1.177	0.769
Science and maths	-0.88 (-3.57, 1.80)	1.364	0.517
Performing arts	0.45 (-3.26, 4.16)	1.887	0.811
Social science	-3.74 (-8.30, 0.83)	2.326	0.108
PE	0.67 (-4.91, 6.24)	2.842	0.814
Other	2.25 (-3.14, 7.65)	2.742	0.412
Class size			
29 and over	Reference		
Under 29	-2.21 (-3.99, -0.43)	0.906	0.015

The data shows that there were only small clustering effects with most of the variables retaining similar statistical significance. All the categories for the variable 'subject', the categories, 'most of the time speaking over background noise', 'most of the time speaking louder than normal', the sex category 'prefer not to say', all smoking categories, asthma, voice training, pupil premium and hours teaching per week all remained insignificant. The only variable to lose its significance when clustering was accounted for was the category sometimes speaking over background noise which went from -3.78 (95% CI -7.44, -0.11) and a p value of 0.044 to -3.46 (95% CI -7.13, 0.22) and a p value of 0.065. When accounting for clustering the WEMWBS and age slightly reduced their

coefficients indicating a small decrease in association with scores on the VoiSS. In the multiple regression before clustering was accounted for the WEMWBS the coefficient was -0.32 (95% CI -0.43, -0.21) and when clustering was accounted for the coefficient became -0.31 (95% CI -0.41, -0.20). Therefore, when clustering was accounted for there was a predicted 0.31 point decrease (SE 0.06) in the VoiSS for every one unit increase in the WEMWBS, with the true estimate ranging from a 0.41 decrease to a 0.20 point decrease in scores on the VoiSS with every point increase on the WEMWBS. The result remained statistically significant with a p value of <0.001. Age reduced it's coefficient from 0.11 (95% CI 0.03, 0.19) to 0.10 (95% CI 0.02, 0.18) when clustering was accounted for with significance reduced to a p value of 0.015 on the multilevel model. GerdQ on the other hand slightly strengthened its association with the VoiSS. The coefficient increased from 1.28 (95%CI 0.87, 1.69) to 1.29 (95% CI 0.87, 1.70) with the association remaining highly statistically significant (p=<0.001).

When observing the categorical variables, the category 'male' showed no clustering effect at all with its coefficients and confidence intervals remaining exactly the same on each model, -3.48 (95%CI -5.59, -1.37) with a p value of 0.001. Some of the categories increased their coefficients when accounting for clustering showing a greater difference between the reference category and the category in question. These included the category 'never' for the variable 'speaking louder than normal' with coefficients increasing from -11.11 (95% CI 15.48, -6.74) in the multivariable model to -11.35 (95%CI -15.73, -6.98) in the multilevel model. The category 'sometimes' for the variable 'speaking louder than normal' increased its coefficient from -8.14 (95%CI -11.17, -5.11) to -8.23 (95% CI -11.26, -5.20) when accounting for clustering. The category, 'yes confirmed by a doctor' for the variable respiratory infection, increased its negative score from -4.03 (95% CI -5.79, -2.27) to -4.15 (95% CI -5.91, -2.39) when accounting for clustering, thus showing that those who had a respiratory infection confirmed by a doctor were predicted to have VoiSS scores on average 4.2 points lower (SE 0.9) than teachers without a respiratory infection. The p value remained at <0.001. The category 'under 29' for the variable 'class size' went from a coefficient of -2.17 (95% CI -3.93, 0.42) to -2.21 (-3.99, -0.45) with the p value retained at 0.015.

The other categories that were found to be statistically significant slightly reduced their coefficients suggesting less difference in VoiSS scores between the reference category and the category of interest. The category, 'never' for the variable 'speaking over background noise' reduced from -5.55 (95% CI 9.33, -1.77) and a p value of 0.004 in the multiple linear regression to a less statistically significant result of -5.39 (95%CI -9.18, -1.61) and a p value 0.005 when accounting for clustering. The category, respiratory infection not confirmed by a doctor reduced from 6.50 (95%CI 4.00, 9.01)

in the multiple linear model to 6.26 (95% CI 3.76, 8.77) for the multilevel model. The p value remained at <0.001.

The intraclass coefficient was calculated for the multilevel model. It was 1.172085e-92, and indicates that the proportion of variance in VoiSS scores explained by school membership, is 1.172085e-92. As the value is very close to zero, this concludes that there is very little evidence of clustering or school effects in the model.

4.6.3.1 Interaction

Table 27 shows the results of adding interaction terms to the model. Interactions were investigated between scores on the WEMWBS and the other variables in the multilevel model. An anova analysis was carried out to test the statistical significance of the interactions. The results all had high p values showing no statistically significant difference between the multilevel model with and without interactions. This demonstrates that the results from the multilevel model without interactions are acceptable.

Table 27: Interaction Terms

Interaction	P value
Warwick Edinburgh: sex	0.91
Warwick Edinburgh: age	0.59
Warwick Edinburgh: pupil premium	0.32
Warwick Edinburgh: GerdQ	0.26
Warwick Edinburgh: smoking	0.62
Warwick Edinburgh: Voice training	0.38
Warwick Edinburgh: subject	0.44
Warwick Edinburgh: talk louder	0.44
Warwick Edinburgh: background noise	0.38
Warwick Edinburgh: respiratory infection	0.69
Warwick Edinburgh: asthma	0.75
Warwick Edinburgh: class size	0.37

4.7 Summary

The results of the multilevel model show that the WEMWBS is significantly negatively associated with scores on the VoiSS. Therefore, an increase in Mental wellbeing is associated with a decrease in voice symptoms. It is not possible in this study to detect the direction of the relationship, so it is unclear whether good mental wellbeing positively affects voice functioning or whether a well-functioning voice positively affects someone's mental state. Age showed a moderate association between voice functioning so that as age increases, voice functioning decreases. This corresponds to literature which outlines age related changes to the larynx which may make it more likely that there would be loss of voice functioning as we age (Mathieson and Greene, 2001). The likelihood of

Gerd was positively associated with scores on the VoiSS so that a greater likelihood of Gerd is associated with poorer voice functioning. This agrees with results from other studies with teachers that reflux was associated with voice problems (Devadas et al., 2017a; Sampaio et al., 2012) .

Men were found to have statistically significantly lower scores on the VoiSS than females. This suggests that females are more at risk of getting voice problems. This is backed up in other risk factor studies with teachers, which have found that females are statistically significantly more likely to have voice problems than males (Assuncao et al., 2012; Roy et al., 2004; van Houtte et al., 2012) . Voice training was not found to be statistically significant. VoiSS scores for those who had had and had not had voice training or voice information were not found to be different. However, the number of participants who reported having had voice training was very low so it may be that there were not enough participants in this category to be able to detect a difference. Those with and without asthma were shown to not have statistically different scores to each other. Again, the proportion of participants with an asthma diagnosis may have been too low to detect a meaningful change.

The variables, speaking against background noise and talking louder than normal suggest that those who always speak against background noise and talk louder than normal may be more at risk for increased voice symptoms than those who never speak against background noise. However, there was not a statistically significant association between number of hours someone taught every week and voice symptoms. This study therefore suggests that it is the intensity of voice use rather than the length of time speaking that effects voice functioning.

An interesting result was that those who reported having a respiratory infection confirmed by the doctor had significantly less vocal symptoms than those without a respiratory infection, whereas those who reported having a respiratory infection not confirmed by a doctor had as expected significantly more voice symptoms than those without a respiratory infection. For those who had their infection confirmed by a doctor it may be that they responded to the treatment from the doctor so that their infection cleared up quickly. It may also be that people who go to the doctor when they have a respiratory infection may take more care of their health including their voice, than those who do not go to the doctor when they have a respiratory infection. These issues will be discussed further in the next chapter.

Chapter 5: Discussion and Conclusions

5.0 Introduction

The study aimed to assess the association between wellbeing and voice functioning in teachers in England. Demographic, health and teaching variables were also included and analysed for their association. A cross-sectional design was chosen as it allows associations between outcome variables and predictor variables to be examined in a time efficient and cost-effective way. The discussion and conclusion chapter will begin by looking at the response to the survey and the biases in the study. It well then discuss the VoiSS scores given by participants and discuss the findings of the multi-level model. The chapter will conclude by considering the limitations of the research, the significance of the findings and their implications for teaching and clinical practice, and recommendations for future research.

5.1 Response to the survey

The number of participants who took part in this study was 1205, of whom 1141 taught in state schools. According to the Department of Education, in 2017, when the data was collected, there were 451,900 teachers working full time in state schools in England (Department for Education, 2018c). This means in state schools there was an approximate response rate of 0.25%. However, it must be noted that although the survey was sent to all schools there was no record of which schools passed on the survey to their teachers and therefore how many teachers had the chance to participate. Therefore, the response rate of teachers who actually received surveys is likely to be higher. A low response rate may have the risk of being unrepresentative of a population. However low response rates do not necessarily mean there is a high bias when respondent characteristics are representative of the population (Dillman, 1991; Krosnick, 1999). To find out whether there was a bias, the study compared teachers in the UK population and the study participants in terms of age, ethnicity, type of school (primary, secondary, special, alternative provision), gender and school region. The study found that there was no bias in terms of age and type of school but showed a bias towards those who were White British/Irish, female, working in the West Midlands and working in the state sector. In particular, Indian/Pakistani/Bangladeshi and Black African teachers and those working in London, the South East, North West, North East and Yorkshire and the Humber were underrepresented.

5.1.1 Gender

A bias towards females corresponds to other studies using online surveys (Aerny-Perreten et al., 2015; Ajaz et al., 2016; Deserno et al., 2017; McKinley et al., 2020; Smith, 2008). There has not been much research undertaken to explain this bias. A study by Saleh and Bista (2017) with graduate

students in the US, suggested that male participants were more likely to respond to a survey if they received a reminder and if the questions were short and precise. In this research, the survey and the reminder email were sent out to schools rather than individuals and therefore it is not possible to determine whether the schools forwarded on the reminder to individual teachers and thus whether this affected response rates. The survey had a significant participant burden, taking on average 20 minutes to complete, which may also have affected response rates from male participants. The proportion of males in the teaching population is 25% whereas in this study males represented 19% of the participants.

5.1.2 Ethnicity

In this research, white ethnicity accounted for 93% of participants when it accounts for 91% of the teaching population. This represented a statistically significant bias. Other research suggests that white ethnicity is over represented in web based survey research (Mette et al., 2016; Sterrett et al., 2017). NHS inpatient surveys also have low response rates among black and minority ethnic groups (Sheldon et al., 2007) which may be due to disengagement (Bowling, 2005). A report by Elam et al. (2001) suggested that lack of engagement in health research in the Black African population in the UK was due to a lack of trust towards the wider British population; a cultural preference for not writing things down as written information may be misused, and a feeling that their contribution to research was not important. Other barriers can be the researchers' own attitudes. Sheikh et al. (2009) found this included prejudices and stereotyping of specific groups and perceiving that engaging minority ethnic groups would require too much time and effort.

Any further research with teachers in England should therefore aim to have a more ethnically representative sample. In order to engage minority ethnic groups in research, the importance of communication with BAME groups and community partnerships has been emphasized (Bonevski et al., 2014; Sills and Desai, 1996) and forming trusting relationships (Rooney et al., 2011). It has also been suggested that surveys should be culturally competent (Sheldon et al., 2007). Cultural competence in research 'is the ability of researchers and research staff to provide high quality research that takes into account the culture and diversity of a population when developing research ideas, conducting research and exploring the applicability of research findings' (Catalyst, 2010, 6). Therefore, any future research with teachers in England should involve teaching groups that represent minority ethnic groups and involve them at the design and pretesting stage of the questionnaire.

5.1.3 Region

The sample for this study was biased towards teachers in the West Midlands. This may be because Birmingham City University (BCU) has links with many schools in the area, sending student teachers

out on placement. Therefore, teachers from this area with personal links to BCU may have been more inclined to answer the survey. Research suggests that if participants have a strong connection to an organisation carrying out research there will be a higher response rate than if the participants have a weak connection (Porter and Whitcomb, 2007). Participants have also been shown to have more willingness to participate in research from an organisation with a strong reputation (Fang et al., 2012) or a nationally recognised university (Pan et al., 2014). It may be that BCU does not have as much prominence as other universities, especially outside of the West Midlands, and may explain less response in other regions of England. For future nationwide studies with teachers, it may be advisable to carry out research across several partnership organisations: for example, universities in different geographical regions.

5.2 Discussion of the Results

5.2.1 VoiSS Scores

The results of the analysis show that the mean total score for the VoiSS was 22.54 and the median was 20. It is not possible to see if these scores were comparable to other school teachers or the general population as no other cross-sectional studies were identified that used the VoiSS as an outcome measure in these populations. However, a study carried out in Brazil with professors in higher education, who may have similar occupational risk factors, had a comparable mean for the VoiSS total scores of 20.51 (Dassie-Leite et al., 2020). Studies that have compared VoiSS scores in dysphonic and vocally healthy populations have found mean values for the total VoiSS score ranging from 6.48 (Moreti et al., 2014) to 10.75 (Contreras et al., 2019) for vocally healthy populations, and 21 (Moreti et al., 2014) to 32 (Contreras et al., 2019) for dysphonic populations. Hence, our study of teachers had a much greater mean score than vocally healthy participants and a similar mean to the dysphonic patients in the Moreti et al. (2014) study. The VoiSS scores in our study therefore suggest that teachers in England are at risk of having voice problems. In order to compare the voice symptoms of teachers to the general population in the UK, it would be a useful addition to the literature for a future study to measure voice symptoms in the general population using the VoiSS.

5.2.2. Wellbeing

The primary aim of the study was to investigate the association between vocal functioning and wellbeing. This study found that there was a statistically significant association between scores on the WEMWBS and the VoiSS, when accounting for demographic, health and teaching factors, concluding that those with higher wellbeing are more likely to have lower VoiSS scores and therefore fewer vocal symptoms. This result is reflected in other studies such as da Rocha and Souza (2013) who found that teachers with common mental disorders had significantly higher scores on the VHI than those without common mental disorders. Other studies with teachers have found a

significantly positive association between current or recent voice problems and mental disorder (da Rocha et al., 2017; de Medeiros et al., 2008), current depressive episode (da Rocha et al., 2015) and anxiety (Moy et al., 2015). A link between psychological factors and current or recent voice disorders has also been found with other populations. Marmor et al. (2016) found a link between recent voice disorders and depression in adults in the USA and Pernambuco et al. (2017) found an association with current voice disorders and anxiety symptoms with older adults living in care homes.

The difficulty with all current research including this study is that the causality and the direction of the association between psychological problems or wellbeing and voice problems cannot be understood. Therefore, it is not known whether it is the voice problem negatively influencing wellbeing or wellbeing influencing the voice. It may be that higher wellbeing means that someone is more likely to be relaxed and therefore will not have so much tension in the muscles around the throat and larynx, so is less likely to misuse their voice. In the health literature, it has been proposed that psychological factors such as stress may increase someone's susceptibility to certain illnesses, for example, inflammatory bowel disease (Oligschlaeger et al., 2019) and migraines (Schramm et al., 2015). Therefore, it is possible to conceive that stress and poor mental health may be a contributory factor to other health issues such as voice problems.

Different biological and psychological mechanisms for the association between psychological factors and voice problems have been proposed. The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) has suggested that functional dysphonia may be an outcome of conversion disorder (American Psychiatric Association, 2013). Conversion disorder is a neurological condition where psychological stressors manifest as physical symptoms without an underlying physical illness (Berger et al., 2020). The DSM-5 states that symptoms can include "weakness or paralysis; abnormal movements…reduced or absent speech volume (dysphonia/aphonia)" (p. 319). Anxiety may lead to tension in the intrinsic and extrinsic laryngeal muscles resulting in a functional voice disorder (Nichol et al., 1993) and stress may cause changes in the autonomic nervous system that result in functional voice disorders (Demmink-Geertman and Dejonckere, 2002). However, there is a lack of empirical evidence to support these theories. Research though does suggest that the onset of voice problems can occur around stressful life events (Baker, 2008; Misono et al., 2019; Seifert, 2005). Furthermore, Gadepalli et al. (2019) found that 66.7% of teachers thought that stress had a negative influence on their voice.

On the other hand, it may be that someone who has voice problems cannot communicate effectively and as a result is not able to take part in their usual activities, which could reduce their wellbeing.

Martinez and Cassol (2015) found that patients had significantly lower anxiety and depression scores after speech and language therapy, indicating that improved voice quality may improve psychological functioning. Furthermore, a qualitative study by Misono et al., 2019 suggested that patients with voice problems experienced lowered self-worth, stress at work, and feelings of hopelessness and anxiety due to their voice problems, although they also found that some of the participants' voice problems also seemed to coincide with stressful events. Therefore, it is likely that there is a bidirectional relationship between psychological factors and voice functioning, so that psychological factors affect the voice and voice problems influence psychological health. To understand the directional nature of wellbeing and voice problems more clearly, a vocal questionnaire such as the VoiSS could be included in large cohort studies similar to the 1970 British Cohort Study or the Millennium Cohort Study. This study was able to demonstrate a statistically significant negative association between well-being and voice problems, but due to the design of the study was not able to provide an understanding of the directional nature between well-being and voice problems.

5.2.3 Gastroesophageal Reflux

This is the first study investigating teachers and voice-related risk factors to use a validated tool to measure gastroesophageal reflux. It found a statistically significant positive relationship between scores on the VoiSS and GerdQ scores, with higher scores on the VoiSS linked to higher scores on the GerdQ, meaning a higher likelihood of having GERD was linked to more voice symptoms. The CI for this study was narrow 1.29 (95% CI 0.87,1.70) indicating good precision with the coefficient estimate. This association is reflected in the literature with other studies finding a positive association between teachers having reflux and voice problems. As well as Charn and Mok (2012) and Sampaio et al. (2012) finding a positive association between current voice problems and acid reflux/heartburn in teachers, Devadas et al. (2017a) found a positive association between teachers' voice problems during their lifetime and acid reflux. Research also suggests that patients with reflux represent a large proportion of voice patients. Koufman et al. (2000) found that approximately 50% of patients with voice problems had GERD and Willems-Bloemer et al. (2000) found 39% of their voice patients had reflux. A meta-analysis by Lechien et al. (2019) showed improvement in voice outcomes for patients with laryngopharyngeal reflux (LPR) when proton pump inhibitors (PPIs) (medication to reduce reflux) were given compared to placebo.

5.2.4 Age

When the univariate analysis was completed, there was no significant association between age and scores on the VoiSS. However, when the multilevel model was carried out and all the other variables were considered, age became significant. As age increases, scores on the VoiSS increase, thus

indicating that as someone gets older, they may be more at risk of getting a voice problem. This result however is not supported by the studies analysed in the literature review, most of which showed no association between current or recent voice problems and age. Only the study with Moy et al. (2015) found that teachers in the 40-49 year age category had greater odds of having a voice problem than teachers who were 50 or over but there was no differences in the odds of having a voice disorder between those of teachers less than 40 and teachers 50 and over. Although an age association was found, it is different from this study. Age was not found to be linked to voice disorders in other cross-sectional studies with different populations. However, these studies all use categories for age rather than age as a continuous measure. Thus, the loss of information in using categories may miss a potential association. In this study, older teachers may have misused their voices for longer, which could explain the results. As will be discussed later, only a very small proportion of teachers had had vocal training or information about their voice, so teachers may not have carried out vocal care or protective vocal techniques during their careers.

Voice problems may increase with age due to structural changes that can result in the deterioration of different aspects of the vocal apparatus (Martins et al., 2014a). This includes the ossification of the laryngeal cartilage (Paulsen et al., 2000) the reduction of properties within the larynx that provide a protective function. For example reduced thickness of the lamina propria of vocal folds and reduced density of epithelial cells (Filho et al., 2003), and a decrease in hyaluronic acid and mucous glands (Ohno et al., 2009; Sato and Hirano, 1997) may make older people more susceptible to getting voice disorders. Therefore, the positive association between age and VoiSS scores in this study may be explained by age related structural changes to the vocal apparatus.

5.2.5 Gender

This study found that female teachers were predicted to have statistically significantly higher scores on the VoiSS than male teachers. This is reflected in studies in the literature that found that female teachers had significantly greater odds or greater prevalence rates than male teachers in having current or recent self-reported voice disorders (Assuncao et al., 2012; Lee et al., 2018; Sampaio et al., 2012). Studies with non-teaching populations also report females having a greater risk of self-reported voice problems. Examples include a general Korean population (Hah et al., 2016) and young adults in the USA (Bainbridge et al., 2017).

The reasons for greater prevalence may be biological or it may be that females are just more likely to report problems than males or perhaps a combination of these factors. There are anatomical differences that may explain the difference. Females have shorter and thinner vocal folds, which result in women having on average a higher fundamental frequency than men do. This means the

vocal folds vibrate more, resulting in more collisions of the vocal folds, which may cause a greater likelihood of damage in females (Hunter et al., 2011). Females also have fewer concentrations of hyaluronic acid, a component that is important for shock absorption and wound repair, than men, especially in the superficial layers of the lamina propria (Ward et al., 2002), again resulting in females being more susceptible to damage of the vocal folds. Furthermore, a recent study by Smith et al. (2019) suggested that females require greater lung pressures to achieve the same sound pressure level (SPL) as men, which may make females more at risk of vocal fatigue and subsequent voice problems. Thus, in this study, it may be that females reported higher scores on the VoiSS than males due to a reporting bias, or it may be due to biological differences. Unfortunately, it is not possible to determine which of these are likely.

5.2.6 Vocal demand

An association was found between scores on the VoiSS and speaking against background noise and speaking louder than normal. This finding is in agreement with studies in the literature review that investigated the relationship between voice disorders and vocal demand. A statistically significant association was found with talking a lot or excessively (Rossi-Barbosa et al., 2016), abusive vocal behaviour (da Rocha et al., 2015), raised voice while teaching (Akinbode et al., 2014), working in another activity with intense use of voice (de Medeiros et al., 2008), speaking frequently against background noise (Lee et al., 2010) professional vocal effort (Sampaio et al., 2012) and voice problems.

Speaking against background noise may explain why teachers are at risk of getting voice problems. The Lombard effect states that a speaker raises their voice as the noise level increases and the ability to hear his own voice decreases (Lane and Tranel, 1971). Studies show that there is a significant increase in the speech level of teachers for every 1-decibel increase in noise level (Bottalico and Astolfi, 2012; Puglisi et al., 2017; Sato and Bradley, 2008). Studies have also shown that vocal loading tasks including those where participants have to speak against background noise, increase speaking effort, vocal tiredness, raise the fundamental frequency and sound pressure levels (Herndon et al., 2019; Sundarrajan et al., 2017; Whitling et al., 2015). Increased vocal effort, tiredness, fundamental frequency and SLP over time may lead to vocal problems. Furthermore, research with teachers suggests that reported voice symptoms such as hoarseness and vocal fatigue is significantly correlated with a teacher's average noise exposure (Kristiansen et al., 2014) and kindergarten teachers consider noise to be one of the biggest factors that is detrimental to their voice (Kankare et al., 2011). The finding of this study therefore supports previous literature and suggests that teachers in England also need to be aware of speaking against background noise and talking louder than their normal volume.

5.2.7 Respiratory problems

This study found respiratory tract infections to be significantly associated with scores on the VoiSS which was supported by other studies with teachers (Akinbode et al., 2014; de Medeiros et al., 2008; Lee et al., 2010) and other populations (Bainbridge et al., 2017; Devadas et al., 2018). Respiratory infections cause inflammation of the airways including the larynx which may make speaking more difficult and can cause hoarseness and breathy voice (Harris, 2020). What is interesting in this study is that teachers who had a respiratory infection confirmed by a doctor had significantly lower predicted VoiSS scores than those without a respiratory infection, which was different from those who had a respiratory infection not confirmed by a doctor who had as expected, higher predicted scores. It may be that the participants who went to the doctors for their respiratory infection received treatment, which improved their symptoms. Moreover, perhaps those who went to the doctors to be diagnosed were different in other ways to patients who did not go to the doctors with a respiratory infection and indeed those who did not have a respiratory infection. Perhaps those who had a diagnosis from the doctor were more likely to be concerned about their health in general, including caring and looking after their voices, better than those who did not go to the doctor with a respiratory infection. Perhaps because of their respiratory infection, they took extra care of their voices, which may have explained the lower predicted VoiSS scores.

5.2.8 Asthma

This study found that there was no association between asthma and scores on the VoiSS. This is not reflected in other cross-sectional studies that looked at current voice disorders and asthma. Lee et al. (2010) and Bainbridge et al. (2017) both found that those with asthma had a higher risk of having a voice disorder than those without asthma. The lack of an association in this study may be to do with the treatment for asthma that the participants were taking which may have been effective at controlling their asthma. Although asthma medication itself, specifically inhaled corticosteroids, has been linked to hoarseness (Buhl, 2006), some formulations may be less likely to produce side effects.

5.2.9 Class size

In this study, teachers who taught in class sizes of 29 and over had significantly higher predicted VoiSS scores than teachers who taught in smaller class sizes. This finding is not replicated in the studies in the literature review where there was no association between class size and current voice problem (Akinbode et al., 2014; da Rocha et al., 2017; da Rocha et al., 2015). Perhaps in our study teachers who teach in class sizes of 29 or over have to raise their voices more whereas in other countries this may not be the case. Cultural differences in teaching styles may account for the differences (Cothran et al., 2005).

5.2.10 Voice training/information about the voice

Our study found no association between voice training or receiving information about the voice and current voice functioning. The majority of participants (68.46%) had not received information about the voice, so there may not have been enough people who had received voice training/information to show an association. However, looking at the results the confidence interval is small which suggests enough power. Studies have been conducted that look at the effects of vocal training with teachers. A review by Hazlett et al. (2011) suggested that there was no conclusive evidence to suggest that vocal training improves voice use in professional voice users. As well as teachers, the review looked at studies with singers, call centre staff, and student teachers. They cited small sample sizes, poorly defined risk variables and lack of RCTs as some of the limitations. Since this review, a study by (Faham et al., 2016) found that teachers undertaking an eight-week voice-training program significantly improved their VHI scores compared to the control group whose VHI scores significantly decreased after 8 weeks. However, this study did not follow up with teachers and measure VHI scores to see if the training program was effective long-term. A study by Nusseck et al. (2019) compared the scores of the VHI-12 with student teachers who had had voice training and controls two years after the training had been completed and found that there was no significant difference in scores between the two groups. This suggests that training is effective in the short term but may not offer long-term benefits. It may be that the teachers in our sample had had training early in their careers or during teacher training and perhaps did not remember to utilise the skills and knowledge they had obtained. However, it is clear, more robust research needs to be carried out to look at the long-term effectiveness of voice training programs.

5.2.11 Smoking

The study did not find a significant association between smoking status and scores on the VoiSS. This is in agreement with other cross sectional studies that included teachers, with no link between smoking and voice disorders being found (de Medeiros et al., 2008; Roy, 2005). However, a study by Byeon (2015) which investigated the association between smoking and dysphonia in the Korean general population found current smokers had a significantly higher risk for current self-perceived voice problems than non-smokers, and heavy smokers had a significantly higher risk than non-smokers. Furthermore, smoking is a major contributory factor to Reinke's oedema (Gugatschka et al., 2019; Tavaluc and Tan-Geller, 2019) and is a significant risk factor for head and neck cancer (Lubin et al., 2007; Maasland et al., 2014; Wyss et al., 2013). The reason no association was found in this study between voice symptoms and smoking may be that there were not enough teachers in the sample to detect a difference. Only 3% (n=36) reported smoking every day and 3% (n=35) reported

smoking occasionally. Another reason may be that teachers who have damaged their voices through smoking, for example those who have developed cancer have left the profession.

5.2.12 Subject taught

No association was found between the subject teachers taught and scores on the VoiSS. Few studies with teachers have investigated whether subject is linked to voice problems, and there are no other studies that show a link between current voice problems and subject taught. In cross- sectional studies with teachers, there has been an association between ever having had a voice disorder and teachers of arts and theatre studies (Charn and Hwei Mok, 2012), teaching chemical sciences (Thibeault et al., 2004), and teaching physical education (PE) (Smith et al., 1998a). Additional research has suggested PE teachers may be more at risk of getting voice problems. A study by Kristiansen et al. (2014) suggested that vocal loudness was higher in sports teachers compared to other subjects, with sports teachers more likely to speak with a loud (86-91Decibels (Db)) or very loud voice (>/= 92Db). This may make PE teachers more likely to develop voice problems. However, this is not something that was indicated in this study. PE teachers though, only accounted for 2% (n=24) of the sample so the numbers are likely to be too small to detect a difference.

5.3 Limitations of the study

There are a few limitations to this study that need to be addressed. The first is that as this was a cross-sectional study, it was unable to establish causality or directionality. Therefore, it is not possible to say whether the risk factors in this study directly caused the levels of voice symptoms in participants. Additionally, it is not possible to say whether levels of wellbeing are affected by voice functioning or voice functioning is affected by wellbeing. In order to help determine directionality, a longitudinal study would need to have been carried out. Unfortunately, due to time and resources, this was not possible. Longitudinal research with voice disorders needs to be undertaken to have a better understanding of the relationship between risk factors and voice problems.

Representation of the total teaching population was not achieved. There was a poor response rate to the study, which may have resulted in sampling bias. However, as it has already been acknowledged it is not possible to determine the true response rate, so the rate of 0.25 may be artificially low. Although school level and age were representative of the population, biases were found for gender, ethnicity and school region. The study did attempt to reduce sampling bias by inviting all schools to participate in the study. However, it is probable that not all head teachers or administrators that received the email decided to forward it on to their staff. The schools and the teachers working in the schools that decided to participate may have been different from those who

did not participate. However, due to the nature of the research, it was not possible to gather information on the characteristics of the non-participating schools and teachers.

Using a census approach for recruitment may have been too ambitious for the scale of the project and in hindsight, it may have been more appropriate to concentrate efforts on the West Midlands. However, the study aimed to get information that would be generalisable to all teachers in England, and just focusing on the West Midlands would have limited generalisability. Sending out emails to all schools in England only managed a small response. As such a large number of emails were being sent, it was not possible to personalise the emails to individual head teachers' names. Lack of personalisation in the email may have meant the emails were not forwarded to the teachers in the school. Research suggests that response rates are better when emails are personalised (Heerwegh, 2005; Sauermann and Roach, 2013; Sinclair et al., 2012). Consequently, after getting a poor response from the mass email, personalised emails were sent to teachers in the West Midlands, whose contact names were obtained from the School of Education at BCU. For future nationwide research with teachers, it would be advisable to collaborate with Education departments at other universities to get contact details for schools, so all emails can be personalised, therefore increasing the likelihood of a response.

As an additional recruitment strategy, the email was advertised on social media, which may have caused a bias in respondents. Those who respond to health surveys on social media may be different from those who respond via other methods. A study that compared the recruitment of participants via Facebook and Instagram with those who were recruited via letter and a follow-up phone call, found those who responded to social media were significantly more likely to be White (Benedict et al., 2019). A study that compared participants who were recruited via Twitter with those recruited via a national survey found that those recruited via Twitter were more likely to be White, female, and under 45 (Keaver et al., 2019). In this study, it was not possible to determine who responded to the survey via social media and therefore if it accounted for any of the biases in the study population. Despite the possibility of bias, using social media meant being able to reach teachers who had not received the questionnaire from their head teachers and therefore was able to help increase the number of responses.

As this study used the VoiSS as a continuous measure, it is difficult to compare the findings with other studies, as most have measured whether a teacher has or does not have a voice problem and used ORs or PRs to measure associations. Additionally, as the VoiSS was used as a continuous measure this study was not able to record the prevalence of voice problems and therefore it is not known how the prevalence of teachers in England compares to other countries. A study in the UK

has measured prevalence for voice problems in teachers in England to be 30%, but their study only included teachers in the North West of England so may not be generalisable to the rest of England (Gadepalli et al., 2019). Additionally, they defined a voice problem as being a VHI-10 total score of more than zero, which may not be a valid way of measuring a voice problem. In order to be able to measure the prevalence of voice problems in teachers in the UK, studies need to be undertaken that create cut off scores for validated scales such as the VoiSS or the VHI with UK populations.

5.4 Recommendations

5.4.1 Further research

This section expands on the recommendations that have already been presented in this chapter. As there is a lack of robust research in the field of risk factors and voice disorders in teachers, especially in the UK, more research of this nature needs to be carried out. Additionally, more research needs to be undertaken to understand the relationship between wellbeing and voice problems in teachers. As well as the need for longitudinal studies, there is a lack of qualitative research with teachers regarding voice problems and wellbeing. Although qualitative research cannot be generalisable to a population, it can give a different perspective to a research problem and has the advantage of being able to gather rich data, which may enhance understanding. As voice problems are multidimensional, it is difficult to get a full understanding through quantitative studies alone. Only using quantitative methodologies may limit the scope of the understanding between voice problems and wellbeing. A qualitative approach has been used in other aspects of speech and language disorders to gain an understanding of patients' lived experience, for example, aphasia (Manning et al., 2019; Simmons-Mackie and Lynch, 2013) and stuttering (Beilby et al., 2013; Daniels et al., 2006; Hearne et al., 2008) and would be a useful addition to the literature for voice problems. A recent qualitative study with voice patients has been carried out in the US (Misono et al., 2019) but a separate study with teachers would be useful to see if they have different experiences compared to voice patients in general. Results from a qualitative study could also perhaps help to inform a questionnaire around wellbeing and voice that could be used in quantitative research.

5.4.2 Recommendations for practice

As the direction of the association between wellbeing and voice functioning is not known, it is recommended that schools help teachers improve both their voice functioning and the wellbeing of their staff. Although this study did not show an association between voice training and voice problems, teachers may have forgotten the skills they were taught or may not use them. Schools need to ensure that all teachers have access to voice training or information on the voice and include it as part of school wellbeing policies and strategies. Teachers' voices should be monitored regularly, perhaps by SLTs working in the schools. Those identified as at risk of getting a voice

problem could be referred for voice training or given voice care information; those with current voice problems could be referred to their GPs. In addition, schools should have strategies for enhancing wellbeing and ensure that these are implemented. The organisation Mentally Healthy Schools provides information and advice for supporting staff wellbeing (Mentally Healthy Schools, 2020).

In clinical practice, clinicians should assess the wellbeing of their voice patients, which can be done informally or through a questionnaire such as the WEMBS. If appropriate, as part of clinical treatment, strategies to improve wellbeing or manage stress can be discussed and if necessary, clients referred on for psychological treatment or counselling. On voice care information sheets given to voice patients and teachers, strategies for improving wellbeing could be included.

In any voice care information given to teachers, the link between reflux and voice problems should be stated so that teachers are aware and can get appropriate treatment. The National Education Union does provide information on Voice Care (National Education Union, 2019a) but unfortunately does not include reflux as a contributory factor. It would be advisable for the NEU and other teaching unions to provide this information.

Teachers who teach in class sizes of 29 and over, and who identify themselves as always or often having to talk louder than normal or against background noise should be given the opportunity to have amplification in their classrooms to lessen their vocal demand, especially those with a current voice problem. Research suggests that the use of amplification in teachers can reduce vocal intensity in those with voice problems (Assad et al., 2019). Additionally, research suggests that amplification may be an effective treatment for voice disorders in teachers (Bovo et al., 2013; Roy et al., 2002). However, there is a lack of research to indicate whether amplification is a useful preventative measure. Research that is more robust needs to be carried out in this field.

5.5 Conclusion

The research set out to investigate the risk factors, primarily wellbeing, associated with voice symptoms in teachers working in England, using a multilevel model. It is the first study to investigate the association of wellbeing with voice symptoms. The study concluded that teachers have more voice symptoms when they have lower wellbeing, are older, have a higher likelihood of having GERD, always speak against background noise, always speak louder than normal, teach in class sizes of 29 or above and have respiratory infections not diagnosed by a doctor. It showed that VoiSS scores in the study population are high compared to vocally healthy people and are comparable with some dysphonic populations showing that teachers in England may be particularly at risk of having voice problems. It has been identified that further studies are needed to be able to better understand the

relationships between risk factors and voice problems. It is suggested that subsequent research should consider undertaking qualitative research to gather information on the subjective experiences of teachers with voice problems, to better understand the unique problems that this population face.

References

Aday, L. A. and Cornelius, L. J. (2006) *Designing and conducting health surveys: a comprehensive guide.* 3rd edn. San Francisco, Calif: Jossey-Bass.

Aerny-Perreten, N., Domínguez-Berjón, M. F., Esteban-Vasallo, M. D. and García-Riolobos, C. (2015) Participation and factors associated with late or non-response to an online survey in primary care. *Journal of evaluation in clinical practice,* 21(4), pp. 688-693.

Aghadoost, O., Moradi, N., Aghadoost, A., Montazeri, A., Soltani, M. and Saffari, A. (2016) A comparative study of Iranian female primary school teachers' quality of life with and without voice complaints. *Journal of Voice,* 30(6), pp. 688-692.

Ajaz, A., David, R., Brown, D., Smuk, M. and Korszun, A. (2016) BASH: badmouthing, attitudes and stigmatisation in healthcare as experienced by medical students. *BJPsych bulletin,* 40(2), pp. 97-102.

Akinbode, R., Lam, K. B. H., Ayres, J. G. and Sadhra, S. (2014) Voice disorders in Nigerian primary school teachers. *Occupational Medicine,* 64(5), pp. 382-386.

Alanazi, R., Alrahim, A., Bayounos, S., Al-Ghuwainem, A. and Al-Bar, M. H. (2018) Association Between Voice Handicap Index and Reflux Symptom Index: A cross-sectional study of undiagnosed general and teacher cohorts in Saudi Arabia. *Sultan Qaboos University Medical Journal,* 18(3), pp. e350-e354.

Albustan, S. A., Marie, B. S., Natour, Y. S. and Darawsheh, W. B. (2018) Kuwaiti teachers' perceptions of voice handicap. *Journal Of Voice,* 32(3), pp. 319-324.

Alin, A. (2010) Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics,* 2(3), pp. 370-374.

Althubaiti, A. (2016) Information bias in health research: definition, pitfalls, and adjustment methods. *Journal of multidisciplinary healthcare,* 9, p. 211.

Altman, D. G. and Royston, P. (2006) The cost of dichotomising continuous variables. *British Medical Journal,* 332(7549), pp. 1080-1080.

Alva, A., Machado, M., Bhojwani, K. and Sreedharan, S. (2017) Study of risk factors for development of voice disorders and its impact on the quality of life of school teachers in Mangalore, India. *Journal Of Clinical And Diagnostic Research: JCDR,* 11(1), pp. MC01-MC05.

American Psychiatric Association (2013) *Diagnostic and statistical manual of mental disorders: DSM-5.* 5th edn. Washington, D.C: American Psychiatric Publishing.

Anand, V., Jones, J. and Gill, P. S. (2015) The relationship between spirituality, health and life satisfaction of undergraduate students in the UK: An online questionnaire study. *Journal of Religion and Health,* 54(1), pp. 160-172.

Arffa, R. E., Krishna, P., Gartner-Schmidt, J. and Rosen, C. A. (2012) Normative values for the Voice Handicap Index-10. *Journal of Voice,* 26(4), pp. 462-465.

Aronson, A. E. and Bless, D. M. (2009) *Clinical voice disorders.* 4th edn. New York: Thieme.

Assad, J. P., Gama, A. C. C., Santos, J. N. and de Castro Magalhães, M. (2019) The Effects of Amplification on vocal dose in teachers with dysphonia. *Journal of Voice,* 33(1), pp. 73-79.

Assuncao, A. A., Bassi, I. B., de Medeiros, A. M., Rodrigues, C. D. and Gama, A. C. C. (2012) Occupational and individual risk factors for dysphonia in teachers. *Occupational Medicine,* 62(7), pp. 553-559.

Bainbridge, K. E., Roy, N., Losonczy, K. G., Hoffman, H. J. and Cohen, S. M. (2017) Voice disorders and associated risk markers among young adults in the United States. *Laryngoscope,* 127(9), pp. 2093-2099.

Baker, J. (2008) The role of psychogenic and psychosocial factors in the development of functional voice disorders. *International Journal of Speech-Language Pathology,* 10(4), pp. 210-230.

Barros, A. J. and Hirakata, V. N. (2003) Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology,* 3(1), p. 21.

Bartlett, J.E., Kotrlik, J.W., and Higgins, C.C. (2001) Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, 19(1), pp.43-50.

Behlau, M., Madazio, G., Moreti, F., Oliveira, G., dos Santos, L. d. M. A., Paulinelli, B. R. and Couto Junior, E. d. B. (2015) Efficiency and cutoff Values of self-assessment instruments on the impact of a voice problem. *Journal of Voice,* 30(4), pp. 506.e509-506.e518.

Behlau, M., Pontes, P. and Moreti, F. (2018) *Higiene vocal: cuidando da voz.* Thieme Revinter Publicações LTDA.

Behlau, M., Zambon, F., Guerrieri, A. C. and Roy, N. (2012) Epidemiology of voice disorders in teachers and nonteachers in Brazil: Prevalence and adverse effects. *Journal of Voice,* 26(5), pp. e9-e18.

Behrens, T., Taeger, D., Wellmann, J. and Keil, U. (2004) Different methods to calculate effect estimates in cross-sectional studies. *Methods of information in medicine,* 43(05), pp. 505-509.

Beilby, J. M., Byrnes, M. L., Meagher, E. L. and Yaruss, J. S. (2013) The impact of stuttering on adults who stutter and their partners. *Journal of Fluency Disorders,* 38(1), pp. 14-29.

Bell, M. L. and Fairclough, D. L. (2014) Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Statistical Methods in Medical Research,* 23(5), pp. 440-459.

Benedict, C., Hahn, A. L., Diefenbach, M. A. and Ford, J. S. (2019) Recruitment via social media: advantages and potential biases. *Digital health,* 5, p. 2055207619867223.

Benninger, M. S., Ahuja, A. S., Gardner, G. and Grywalski, C. (1998) Assessing outcomes for dysphonic patients. *Journal of Voice,* 12(4), pp. 540-550.

Berger, F., Zieve, D. and Conaway, B. (2020) *Conversion Disorder.* Web page. Available at: https://medlineplus.gov/ency/article/000954.htm [Accessed 23rd June 2020].

Berry, W. D., Feldman, S. and Stanley Feldman, D. (1985) *Multiple regression in practice.* Sage.

Bhaskaran, K. and Smeeth, L. (2014) What is the difference between missing completely at random and missing at random? *International Journal of Epidemiology,* 43(4), pp. 1336-1339.

Bland, J. M. (2009) The tyranny of power: is there a better way to calculate sample size? *British Medical Journal,* 339.

Bland, J. M. and Altman, D. G. (1997) Cronbach's alpha. *British Medical Journal,* 314(7080), pp. 572-572.

Bonevski, B., Randell, M., Paul, C., Chapman, K., Twyman, L., Bryant, J., Brozek, I. and Hughes, C. (2014) Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Medical Research Methodology,* 14(1), p. 42.

Bottalico, P. and Astolfi, A. (2012) Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms. *Journal of the Acoustical Society of America,* 131(4), pp. 2817-2827.

Bovo, R., Trevisi, P., Emanuelli, E. and Martini, A. (2013) Voice amplification for primary school teachers with voice disorders: A randomized clinical trial. *International Journal of Occupational Medicine and Environmental Health,* 26(3), pp. 363-372.

Bowling, A. (2005) Mode of questionnaire administration can have serious effects on data quality. *Journal of public health,* 27(3), pp. 281-291.

Braunholtz, S., Davidson, S., Myant, K. and O'Connor, R. (2007) *Well? What do you think?(2006): The third national Scottish survey of public attitudes to mental health, mental wellbeing and mental health problems.* Scottish Government Edinburgh, Scotland.

Bruce, P., Bruce, A. and Gedeck, P., 2020. *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python*. 2nd edn. Sebastopol, CA: O'Reilly Media.

Buhl, R. (2006) Local oropharyngeal side effects of inhaled corticosteroids in patients with asthma. *Allergy,* 61(5), pp. 518-526.

Buxton, R. (2008) *Statistics: Multilevel Modelling*. Conference Proceedings

Byeon, H. (2015) The association between lifetime cigarette smoking and dysphonia in the Korean general population: findings from a national survey. *PeerJ,* 3, p. e912.
Byeon, H. (2019) The Risk Factors Related to Voice Disorder in Teachers: A Systematic Review and Meta-Analysis. *International Journal Of Environmental Research And Public Health,* 16(19).
Campbell, S. M. and Roland, M. O. (1996) Why do people consult the doctor? *Family practice,* 13(1), pp. 75-83.
Cannes do Nascimento, N., Dos Santos, A. P., Sivasankar, M. P. and Cox, A. (2020) Unraveling the molecular pathobiology of vocal fold systemic dehydration using an in vivo rabbit model. *PLOS ONE,* 15(7), p. e0236348.
Carding, P. N., Horsley, I. A. and Docherty, G. J. (1999) A study of the effectiveness of voice therapy in the treatment of 45 patients with nonorganic dysphonia. *Journal of Voice,* 13(1), pp. 72-104.
Care Quality Commission (2016) *NHS Patient Surveys. Response Rates for the community mental health survey.* Online publication. Available at: http://www.nhssurveys.org/Filestore/Community_mental_health_response_rates.pdf [Accessed 9th January 2020].
Care Quality Commission (2020) *Maternity Survey, 2018*. [data collection]. UK Data Service. SN: 8596. Available at: http://doi.org/10.5255/UKDA-SN-8596-1
Catalyst, H. (2010) Cultural competence in research. *Program for Faculty Development & Diversity. Monitoring, Evaluation, Accountability and Learning (MEAL).* The Open University. USA.
Chan, R. W. and Tayama, N. (2002) Biomechanical effects of hydration in vocal fold tissues. *Otolaryngology-Head and Neck Surgery,* 126(5), pp. 528-537.
Charn, T. C. and Hwei Mok, P. K. (2012) Voice problems amongst primary school teachers in Singapore. *Journal of Voice,* 26(4), pp. e141-147.
Chen, S. H., Chiang, S.-C., Chung, Y.-M., Hsiao, L.-C. and Hsiao, T.-Y. (2010) Risk factors and effects of voice problems for teachers. *Journal of Voice,* 24(2), pp. 183-192.
Chowdhury, M. Z. I. and Turin, T. C. (2020) Variable selection strategies and its importance in clinical prediction modelling. *Family Medicine and Community Health,* 8(1).
Chui, J. C. H. and Ma, E. P. M. (2018) The impact of dysphonic voices on children's comprehension of spoken language. *Journal of Voice*.
Contreras, F., Moreti, F., Vivero, M., Malebran, C. and Behlau, M. (2019) Cross-cultural adaptation, validation, and cutoff values of the Chilean version of the Voice Symptom Scale: VoiSS. *Journal of Voice*. S0892-1997(19)30207-3. doi: 10.1016/j.jvoice.2019.09.020.
Cook, C., Heath, F. and Thompson, R. L. (2000) A meta-analysis of response rates in web-or internet-based surveys. *Educational and psychological measurement,* 60(6), pp. 821-836.
Cook, R. D. (1977) Detection of influential observation in linear regression. *Technometrics,* 19(1), pp. 15-18.
Cothran, D. J., Kulinna, P. H., Banville, D., Choi, E., Amade-Escot, C., MacPhail, A., Macdonald, D., Richard, J.-F., Sarmento, P. and Kirk, D. (2005) A cross-cultural investigation of the use of teaching styles. *Research quarterly for exercise and sport,* 76(2), pp. 193-201.
Creswell, J. W. and Creswell, J. D. (2018) *Research design: qualitative, quantitative & mixed methods approaches.* 5th;International student; edn. Los Angeles: SAGE.
Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika,* 16(3), pp. 297-334.
Cumming, G. and Finch, S. (2005) Inference by eye: confidence intervals and how to read pictures of data. *American psychologist,* 60(2), p. 170.
Cunningham, C. T., Quan, H., Hemmelgarn, B., Noseworthy, T., Beck, C. A., Dixon, E., Samuel, S., Ghali, W. A., Sykes, L. L. and Jetté, N. (2015) Exploring physician specialist response rates to web-based surveys. *BMC Medical Research Methodology,* 15(1), pp. 1-8.

Cutiva, L. C. and Burdorf, A. (2016) Work-related determinants of voice complaints among school workers: An eleven-month follow-up study. *American Journal of Speech-Language Pathology,* 25(4), pp. 590-597.
Cutiva, L. C. C. and Burdorf, A. (2014) Factors associated with voice-related quality of life among teachers with voice complaints. *Journal of Communication Disorders,* 52, pp. 134-142.
Cutiva, L. C. C., Vogel, I. and Burdorf, A. (2013) Voice disorders in teachers and their associations with work-related factors: A systematic review. *Journal of Communication Disorders,* 46(2), pp. 143-155.
da Rocha, L. M., Bach, S. D., do Amaral, P. L., Behlau, M. and Souza, L. D. D. (2017) Risk factors for the incidence of perceived voice disorders in elementary and middle school teachers. *Journal of Voice,* 31(2), pp. 258-e212.
da Rocha, L. M., Behlau, M. and Souza, L. D. D. (2015) Behavioral dysphonia and depression in elementary school teachers. *Journal of Voice,* 29(6), pp. 712-717.
da Rocha, L. M. and Souza, L. D. D. (2013) Voice Handicap Index associated with common mental disorders in elementary school teachers. *Journal of Voice,* 27(5), pp. 595-602.
Daniels, D. E., Hagstrom, F. and Gabel, R. M. (2006) A qualitative study of how African American men who stutter attribute meaning to identity and life choices. *Journal of Fluency Disorders,* 31(3), pp. 200-215.
Dassie-Leite, A. P., Cercal, G. C. S., de Paula, A. L., Novis, J. M. M. and Ribeiro, V. V. (2020) Vocal symptoms in Brazilian professors: Self-perception and relationship factors. *Journal of Voice*.
de Alvear, R. M. B., Barón, F. J. and Martínez-Arquero, A. G. (2011) School teachers' vocal use, risk factors, and voice disorder prevalence: Guidelines to detect teachers with current voice problems. *Folia Phoniatrica et Logopaedica,* 63(4), pp. 209-215.
de Jong, F., Kooijman, P. G. C., Thomas, G., Huinck, W. J., Graamans, K. and Schutte, H. K. (2006) Epidemiology of voice problems in Dutch teachers. *Folia Phoniatrica et Logopaedica,* 58(3), pp. 186-198.
De Leeuw, J., Meijer, E. and Goldstein, H. (2008) *Handbook of multilevel analysis.* New York: Springer.
de Medeiros, A. M., Assunção, A. Á. and Barreto, S. M. (2012) Absenteeism due to voice disorders in female teachers: a public health problem. *International Archives of Occupational and Environmental Health,* 85(8), pp. 853-864.
de Medeiros, A. M., Barreto, S. M. and Assunção, A. Á. (2008) Voice disorders (dysphonia) in public school female teachers working in Belo Horizonte: Prevalence and associated factors. *Journal of Voice,* 22(6), pp. 676-687.
Deary, I. J., Scott, S., Wilson, I. M., White, A., MacKenzie, K. and Wilson, J. A. (1997) Personality and psychological distress in dysphonia. *British Journal of Health Psychology,* 2(4), pp. 333-341.
Deary, I. J., Wilson, J. A., Carding, P. N. and MacKenzie, K. (2003) VoiSS: A patient-derived Voice Symptom Scale. *Journal of Psychosomatic Research,* 54(5), pp. 483-489.
Demmink-Geertman, L. and Dejonckere, P. H. (2002) Nonorganic habitual dysphonia and autonomic dysfunction. *Journal of Voice,* 16(4), pp. 549-559.
Department for Communities and Local Government (2017) *English Housing Survey, 2015-2016: Household Data*. [data collection]. UK Data Service. SN: 8185. Available at: http://doi.org/10.5255/UKDA-SN-8185-1 [Accessed 1st April 2018]
Department for Education (2017) *Schools in England*. Excel spreadsheet. Available at: https://www.gov.uk/government/publications/schools-in-england [Accessed 1st April 2018].
Department for Education (2018a) *Pupil premium 2017 to 2018: allocations (school level)*. Excel spreadsheet. Available at: https://www.gov.uk/government/publications/pupil-premium-conditions-of-grant-2017-to-2018 [Accessed 1st April 2018].
Department for Education (2018b) *School Workforce Census 2017. November 2017*. Excel spreadsheet. Available at: https://www.gov.uk/government/statistics/school-workforce-in-england-november-2017 [Accessed 12th July 2019].

Department for Education (2018c) *School Workforce in England: November 2017.* Excel spreadsheet. Available at: https://www.gov.uk/government/statistics/school-workforce-in-england-november-2017 [Accessed 8th July 2019].
Department for Education (2020) *Pupil Premium.* Web page. London: Department for Education. Available at: https://www.gov.uk/government/publications/pupil-premium/pupil-premium [Accessed 10th February 2020].
Deserno, M. K., Borsboom, D., Begeer, S. and Geurts, H. M. (2017) Multicausal systems ask for multicausal approaches: A network perspective on subjective well-being in individuals with autism spectrum disorder. *Autism,* 21(8), pp. 960-971.
Devadas, U., Bellur, R. and Maruthy, S. (2017a) Prevalence and risk factors of voice problems among primary school teachers in India. *Journal of Voice,* 31(1), pp. 117-117.e110.
Devadas, U., Hegde, M. and Maruthy, S. (2017b) Prevalence and risk factors of self-reported voice problems among Yakshagana artists. *Journal of Voice*, 33(1), pp. 124.e35-124.e47
Devadas, U., Hegde, M. and Maruthy, S. (2018) Prevalence of and risk factors for self-reported voice problems among Hindu temple priests. *Journal of voice*, 33(5), pp. 805.e1-805.e12
Devadas, U., Jose, N. and Gunjawate, D. (2015) Prevalence and influencing risk factors of voice problems in priests in Kerala. *Journal of Voice,* 30(6), pp. 771.e727-771.e732.
Dewey, J. (2008) *The Middle Works of John Dewey, Volume 14, 1899-1924: Human Nature and Conduct 1922.* Illinois: Southern Illinois University Press.
Dewey, J. (2018) *Logic-The theory of inquiry.* Redditch: Read Books Ltd.
Dillman, D. A. (1991) The design and administration of mail surveys. *Annual Review of Sociology,* 17(1), pp. 225-249.
Dillman, D. A., Smyth, J. D. and Christian, L. M. (2014) *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, 4th Edition.* Indianapolis: John Wiley & Sons, Inc.
Dogan, M., Eryuksel, E., Kocak, I., Celikel, T. and Sehitoglu, M. A. (2007) Subjective and objective evaluation of voice quality in patients with asthma. *Journal Of Voice,* 21(2), pp. 224-230.
Education Scotland (2018) *Scottish Education System.* Web page. Available at: https://education.gov.scot/scottish-education-system [Accessed March 22nd 2018].
Education Support (2020). Web page. Available at: https://www.educationsupport.org.uk [Accessed 9th November 2020].
Elam, G., McMunn, A. and Nazroo, J. (2001) *Feasibility study for health surveys among black African people living in England: Final Report--implications for the Health Surveys for England.* London: Joint Health Surveys Unit National Centre for Social Research
Emanuel, E. J., Wendler, D. and Grady, C. (2000) What makes clinical research ethical? *Journal of the American Medical Association,* 283(20), pp. 2701-2711.
Erickson, E. and Sivasankar, M. (2010) Evidence for adverse phonatory change following an inhaled combination treatment. *Journal Of Speech, Language, And Hearing Research: JSLHR,* 53(1), pp. 75-83.
Evans, R., Brockman, R., Grey, J., Bell, S., Harding, S., Gunnell, D., Campbell, R., Murphy, S., Ford, T. and Hollingworth, W. (2018) A cluster randomised controlled trial of the Wellbeing in Secondary Education (WISE) Project–an intervention to improve the mental health support and training available to secondary school teachers: protocol for an integrated process evaluation. *Trials,* 19(1), p. 270.
Eysenbach, G. (2004) Improving the quality of web surveys: The checklist for reporting results of Internet e-surveys (CHERRIES). *Journal of Medical Internet Research,* 6(3), pp. 12-16.
Faham, M., Ahmadi, A., Drinnan, M., Saadatmand, N., Fatahi, E. and Jalalipour, M. (2016) The effects of a voice education program on VHI scores of elementary school teachers. *Journal of Voice,* 30(6), pp. 755.e751-755.e711.
Fang, J., Wen, C. and Pavur, R. (2012) Participation willingness in web surveys: exploring effect of sponsoring corporation's and survey provider's reputation. *Cyberpsychology, Behavior, and Social Networking,* 15(4), pp. 195-199.

Fat, L. N., Scholes, S., Boniface, S., Mindell, J. and Stewart-Brown, S. (2017) Evaluating and establishing national norms for mental wellbeing using the short Warwick–Edinburgh mental well-being scale (SWEMWBS): findings from the health survey for England. *Quality of Life Research,* 26(5), pp. 1129-1144.

Fellman, D. and Simberg, S. (2017) Prevalence and risk factors for voice problems among soccer coaches. *Journal of Voice,* 31(1), pp. 121. e129-121. e115.

Ferreira, L. P., de Oliveira Latorre, M. d. R. D., Pinto Giannini, S. P., de Assis Moura Ghirardi, A. C., de Fraga e Karmann, D., Silva, E. E. and Figueira, S. (2010) Influence of abusive vocal habits, hydration, mastication, and sleep in the occurrence of vocal symptoms in teachers. *Journal of Voice,* 24(1), pp. 86-92.

Field, A. P. (2013) *Discovering statistics using IBM SPSS statistics: and sex and drugs and rock 'n' roll.* 4th edn. London: SAGE.

Filho, J. A. X., do Nascimento, P. H. S., Tsuji, D. H. and Sennes, L. U. (2003) Histologic changes in human vocal folds correlated with aging: a histomorphometric study. *Annals of Otology, Rhinology & Laryngology,* 112(10), pp. 894-898.

Finkelhor, B. K., Titze, I. R. and Durham, P. L. (1988) The effect of viscosity changes in the vocal folds on the range of oscillation. *Journal of Voice,* 1(4), pp. 320-325.

Flom, P. L. and Cassell, D. L. (2007) *Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use*. North East SAS Users Group Inc 20th Annual Conference: 11-14th November 2007. Baltimore, Maryland

Fontan, L., Fraval, M., Michon, A., Déjean, S. and Welby-Gieusse, M. (2016) Vocal problems in sports and fitness instructors: A study of prevalence, risk Factors, and need for prevention in France. *Journal of Voice,* 31(2), pp. 261.e233-261.e238.

Freedman, D.A., 2006. On the so-called "Huber sandwich estimator" and "robust standard errors". *The American Statistician*, 60(4), pp.299-302.

Fricker, R. D., Fricker, R. D., Schonlau, M. and Schonlau, M. (2002) Advantages and disadvantages of internet research surveys: Evidence from the literature. *Field Methods,* 14(4), pp. 347-367.

Fuller, E., Mindell, J. and Prior, G. (2018) *Health Survey for England 2017*. London.

Gadepalli, C., Fullwood, C., Ascott, F. and Homer, J. J. (2019) Voice burden in teachers and non-teachers in a UK population: A questionnaire-based survey. *Clinical Otolaryngology,* 44(6), pp. 1045-1058.

Galbraith, S., Daniel, J.A. and Vissel, B., (2010) A study of clustered data and approaches to its analysis. *Journal of Neuroscience*, *30*(32), pp.10601-10608.

Gassull, C., Casanova, C., Botey, Q. and Amador, M. (2010) The impact of the reactivity to stress in teachers with voice problems. *Folia Phoniatrica et Logopaedica,* 62(1-2), pp. 35-39.

Gill, G. A., Johnston, N., Buda, A., Pignatelli, M., Pearson, J., Dettmar, P. W. and Koufman, J. (2005) Laryngeal epithelial defenses against laryngopharyngeal reflux: Investigations of E-cadherin, carbonic anhydrase isoenzyme III, and pepsin. *Annals of Otology, Rhinology and Laryngology,* 114(12 I), pp. 913-921.

Glogowska, M. (2015) Paradigms, pragmatism and possibilities: mixed-methods research in speech and language therapy. *International Journal of Language & Communication Disorders*, pp. 1-10.

Goldberg, D. P. and Williams, P. (1988) *A user's guide to the general health questionnaire.* London: GL Assessment.

Gordis, L. (2014) *Epidemiology.* Fifth edn. Philadelphia, PA: Saunders.

Gosling, R., Bassett, C., Gilby, N., Angle, H. and Catto, S. (2007) Health Education Population Survey: Update from 2006 Survey. *Health Scotland: Glasgow*.

Gray, D. E. (2013) *Doing research in the real world.* London: Sage.

Groves, R. M. (2009) *Survey methodology.* 2nd edn. Hoboken, NJ: John Wiley & Sons.

Gugatschka, M., Darnhofer, B., Grossmann, T., Schittmayer, M., Hortobagyi, D., Kirsch, A., Karpf, E., Brcic, L., Birner-Gruenberger, R. and Karbiener, M. (2019) Proteomic analysis of vocal fold

fibroblasts exposed to cigarette smoke extract: Exploring the pathophysiology of Reinke's Edema. *Molecular & Cellular Proteomics,* 18(8), pp. 1511-1525.
Gurbuzler, L., Inanir, A., Yelken, K., Koc, S., Eyibilen, A. and Uysal, I. O. (2013) Voice disorder in patients with Fibromyalgia. *Auris Nasus Larynx,* 40(6), pp. 554-557.
Hagger-Johnson, G., 2014. *Introduction to research methods and data analysis in the health sciences*. Oxon: Routledge.
Hah, J. H., An, S.-Y., Sim, S., Kim, S. Y., Oh, D. J., Park, B., Kim, S.-G. and Choi, H. G. (2016) A population-based study on the association between rheumatoid arthritis and voice problems. *Clinical Rheumatology,* 35(7), pp. 1873-1878.
Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E. (2014) *Multivariate data analysis.* Seventh Pearson new international edn. Harlow: Pearson Education.
Harrell Jr, F. E. (2015) *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis.* New York: Springer.
Harris, S. (2020) *Coping with colds: upper respiratory infections and the voice.* Web page. Available at: https://www.britishvoiceassociation.org.uk/voicecare_coping-with-colds.htm [Accessed 26th June 2020].
Hazlett, D. E., Duffy, O. M. and Moorhead, S. A. (2011) Review of the Impact of Voice Training on the Vocal Quality of Professional Voice Users: Implications for Vocal Health and Recommendations for Further Research. *Journal of Voice,* 25(2), pp. 181-191.
Hazra, A. (2017) Using the confidence interval confidently. *Journal of thoracic disease,* 9(10), pp. 4125-4129.
Heads Together (2018) *Mentally Healthy Schools.* Web page. Available at: https://www.mentallyhealthyschools.org.uk/?utm_source=headstogether.org.uk&utm_medium=referral [Accessed 26th July].
Hearne, A., Packman, A., Onslow, M. and Quine, S. (2008) Stuttering and its treatment in adolescence: The perceptions of people who stutter. *Journal of Fluency Disorders,* 33(2), pp. 81-98.
Heerwegh, D. (2005) Effects of personal salutations in e-mail invitations to participate in a web survey. *Public opinion quarterly,* 69(4), pp. 588-598.
Hemler, R. J. B., Wieneke, G. H., Labacq, J. and Dejonckere, P. H. (2001) Laryngeal mucosa elasticity and viscosity in high and low relative air humidity. *European Archives of Oto-Rhino-Laryngology,* 258(3), pp. 125-129.
Herndon, N. E., Sundarrajan, A., Sivasankar, M. P. and Huber, J. E. (2019) Respiratory and laryngeal function in teachers: Pre- and postvocal loading challenge. *Journal Of Voice,* 33(3), pp. 302-309.
Hilton, C. E. (2017) The importance of pretesting questionnaires: a field research example of cognitive pretesting the Exercise referral Quality of Life Scale (ER-QLS). *International Journal of Social Research Methodology,* 20(1), pp. 21-34.
Hogikyan, N. D. and Sethuraman, G. (1999) Validation of an instrument to measure voice-related quality of life (V-RQOL). *Journal of Voice,* 13(4), pp. 557-569.
Holmes, S. (2009) Methodological and ethical considerations in designing an Internet study of quality of life: A discussion paper. *International Journal of Nursing Studies,* 46(3), pp. 394-405.
Hosmer Jr, D. W., Lemeshow, S. and Sturdivant, R. X. (2013) *Applied logistic regression.* Chichester: John Wiley & Sons.
Hunter, E. J., Tanner, K. and Smith, M. E. (2011) Gender differences affecting vocal health of women in vocally demanding careers. *Logopedics Phoniatrics Vocology,* 36(3), pp. 128-136.
Huppert, F. A. and Whittington, J. E. (2003) Evidence for the independence of positive and negative well-being: Implications for quality of life assessment. *British Journal of Health Psychology,* 8(1), pp. 107-122.

Hutchings, M., James, K., Maylor, U., Menter, I. and Smart, S. (2006) The recruitment, deployment and management of supply teachers in England. *London: Department for Education and Skills.*

ICO (2018) *Guide to the General Data Protection Regulation* Information Commissioner's Office.

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S. and Newman, C. W. (1997) The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech-Language Pathology,* 6(3), pp. 66-69.

Jackson, C.A. and Smith, J.E. (2003). Study Design. In Gao-Smith, F. and Smith, J.E. (eds) *Key Topics in Clinical Research*. Oxford: BIOS Scientific Publishers, pp. 57-61

Jager, J., Putnick, D. L. and Bornstein, M. H. (2017) II. More than just convenient: The scientific merits of homogeneous convenience samples. *Monographs of the Society for Research in Child Development,* 82(2), pp. 13-30.

Jewell, N. P. (2004) *Statistics for epidemiology.* Boca Raton: Chapman & Hall/CRC.

Jiang, J., Verdolini, K., Aquino, B., Ng, J. and Hanson, D. (2000) Effects of dehydration on phonation in excised canine larynges. *Annals of Otology Rhinology and Laryngology,* 109(6), pp. 568-575.

Johnston, N., Bulmer, D., Gill, G. A., Panetti, M., Ross, P. E., Pearson, J. P., Pignatelli, M., Axford, S. E., Dettmar, P. W. and Koufman, J. A. (2003) Cell biology of laryngeal epithelial defenses in health and disease: further studies. *The Annals of Otology, Rhinology, and Laryngology,* 112(6), pp. 481-491.

Johnston, N., Knight, J., Dettmar, P. W., Lively, M. O. and Koufman, J. (2004) Pepsin and carbonic anhydrase isoenzyme III as diagnostic markers for laryngopharyngeal reflux disease. *Laryngoscope,* 114(12), pp. 2129-2134.

Johnston, N., Wells, C. W., Blumin, J. H., Toohill, R. J. and Merati, A. L. (2007) Receptor-mediated uptake of pepsin by laryngeal epithelial cells. *Annals of Otology, Rhinology and Laryngology,* 116(12), pp. 934-938.

Jones, C. (2011) *Ethical issues in online research.* British Educational Research Association on-line resource. Available at https://www.bera.ac.uk/publication/ethical-issues-in-online-research [Accessed November 23rd 2019]

Jones, R., Junghard, O., Dent, J., Vakil, N., Halling, K., Wernersson, B. and Lind, T. (2009) Development of the GerdQ, a tool for the diagnosis and management of gastro-oesophageal reflux disease in primary care. *Alimentary Pharmacology and Therapeutics,* 30(10), pp. 1030-1038.

Jones, W. P. and Loe, S. A. (2013) Optimal number of questionnaire response categories: More may not be better. *SAGE Open,* 3(2), pp. 1-10.

Kang, H. (2013) The prevention and handling of the missing data. *Korean Journal of Anesthesiology,* 64(5), pp. 402-406.

Kankare, E., Geneid, A., Laukkanen, A. M. and Vilkman, E. (2011) Subjective evaluation of voice and working conditions and phoniatric examination in kindergarten teachers. *Folia Phoniatrica et Logopaedica,* 64(1), pp. 12-19.

Kaushik, V. and Walsh, C. A. (2019) Pragmatism as a research paradigm and its implications for social work research. *Social Sciences,* 8(9), p. 255.

Keaver, L., McGough, A., Du, M., Chang, W., Chomitz, V., Allen, J. D., Attai, D. J., Gualtieri, L. and Zhang, F. F. (2019) Potential of using Twitter to recruit cancer survivors and their willingness to participate in nutrition research and web-based interventions: a cross-sectional study. *JMIR cancer,* 5(1), p. e7850.

Kelley, K. and Bolin, J.H., 2013. Multiple regression. In: T.Teo, ed. *Handbook of quantitative methods for educational research*, Rotterdam: Sense Publishers, pp. 69-101.

Kendall, J. M. (2003) Designing a research project: randomised controlled trials and their principles. *Emercency Medicine Journal,* 20(2), pp. 164-168.

Keyes, C. L., Dhingra, S. S. and Simoes, E. J. (2010) Change in level of positive mental health as a predictor of future risk of mental illness. *American Journal of Public Health,* 100(12), pp. 2366-2371.
Kidger, J., Brockman, R., Tilling, K., Campbell, R., Ford, T., Araya, R., King, M. and Gunnell, D. (2015) Teachers' wellbeing and depressive symptoms, and associated risk factors: A large cross sectional study in English secondary schools. *Journal of Affective Disorders,* 192, pp. 76-82.
Kim, K. H., Kim, R. B., Hwang, D. U., Won, S. J. and Woo, S. H. (2015) Prevalence of and sociodemographic factors related to voice disorders in South Korea. *Journal of Voice,* 30(2), pp. 246e241-246e247.
Kim, S. Y., Lee, J., Choi, Y. S., Kim, J. W., Kwok, S. K., Park, Y. H., Sun, D. I. and Park, S. H. (2018) Do I sound dry? Comparative voice analysis of primary Sjogren's syndrome. *Clinical and Experimental Rheumatology,* 36(3), pp. S130-S136.
Kirkby, H. M., Wilson, S., Calvert, M. and Draper, H. (2011) Using e-mail recruitment and an online questionnaire to establish effect size: A worked example. *BMC Medical Research Methodology,* 11(1), pp. 89-89.
Kirkwood, B. R. and Sterne, J. A. C. (2003) *Essential medical statistics.* 2nd edn. Oxford: Blackwell Science.
Kock, N., Lynn, G. S., Texas, A., University, M. I. and Stevens Institute of, T. (2012) Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association of Information Systems,* 13(7), pp. 546-580.
Kontopantelis, E., Sperrin, M., Mamas, M.A. and Buchan, I.E., 2018. Investigating heterogeneity of effects and associations using interaction terms. *Journal of Clinical Epidemiology*, 93, pp.79-83.
Kosztyła-Hojna, B., Łobaczuk-Sitnik, A., Biszewska, J., Moskal-Jasińska, D., Kraszewska, A., Zdrojkowski, M. and Duchnowska, E. (2018) Subjective and objective assessment of voice quality in pregnancy. *Otolaryngologia polska= The Polish otolaryngology,* 73(2), pp. 1-5.
Koufman, J. A., Amin, M. R. and Panetti, M. (2000) Prevalence of reflux in 113 consecutive patients with laryngeal and voice disorders. *Otolaryngology - Head and Neck Surgery,* 123(4), pp. 385-388.
Krischke, S., Weigelt, S., Hoppe, U., Köllner, V., Klotz, M., Eysholdt, U. and Rosanowski, F. (2005) Quality of life in dysphonic patients. *Journal of Voice,* 19(1), pp. 132-137.
Kristiansen, J., Lund, S. P., Persson, R., Shibuya, H., Nielsen, P. M., Scholz, M., Institutionen för, p., Lund, U., Department of, P. and Lunds, u. (2014) A study of classroom acoustics and school teachers' noise exposure, voice load and speaking time during teaching, and the effects on vocal and mental fatigue development. *International Archives of Occupational and Environmental Health,* 87(8), pp. 851-860.
Krosnick, J. A. (1999) Survey research. *Annual review of psychology,* 50(1), pp. 537-567.
Krouse, J. H., Dworkin, J. P., Carron, M. A. and Stachler, R. J. (2008) Baseline laryngeal effects among individuals with dust mite allergy. *Otolaryngology - Head and Neck Surgery,* 139(1), pp. 149-151.
Kyriakou, K., Petinou, K. and Phinikettos, I. (2018) Risk factors for voice disorders in university professors in Cyprus. *Journal of Voice,* 32(5), pp. 643. e641-643. e649.
Laerd Statistics (2015) *Multiple Regression Using SPSS*. Online statistical tutorials and software guide. Available at: https://statistics.laerd.com/ [Accessed 29th January 2020].
Lane, H. and Tranel, B. (1971) The Lombard sign and the role of hearing in speech. *Journal of speech and hearing research,* 14, pp. 667-709.
Lwanga, S.K., Lemeshow, S. and World Health Organization. (1991) Sample size determination in health studies: a practical manual. *World Health Organisation.*
Lazarus, C. L. (2009) Effects of chemoradiotherapy on voice and swallowing. *Current Opinion in Otolaryngology & Head and Neck Surgery,* 17(3), p. 172.

Lechien, J. R., Finck, C., Huet, K., Fourneau, A.-F., Khalife, M., Delvaux, V., Piccaluga, M., Harmegnies, B. and Saussez, S. (2020) Voice quality as therapeutic outcome in laryngopharyngeal reflux disease: A prospective cohort study. *Journal of Voice,* 34(1), pp. 112-120.

Lechien, J. R., Saussez, S., Schindler, A., Karkos, P. D., Hamdan, A. L., Harmegnies, B., De Marrez, L. G., Finck, C., Journe, F. and Paesmans, M. (2019) Clinical outcomes of laryngopharyngeal reflux treatment: A systematic review and meta-analysis. *The Laryngoscope,* 129(5), pp. 1174-1187.

Lee, S. Y.-y., Lao, X. Q. and Yu, I. T.-s. (2010) A cross-sectional survey of voice disorders among primary school teachers in Hong Kong. *Journal Of Occupational Health,* 52(6), pp. 344-352.

Lee, Y. W., Kim, G. H., Bae, I. H., Park, H. J., Wang, S. G. and Kwon, S. B. (2018) The cut-off analysis using visual analogue scale and cepstral assessments on severity of voice disorder. *Logopedics Phoniatrics Vocology,* 43(4), pp. 175-180.

Levendoski, E. E., Sundarrajan, A. and Sivasankar, M. P. (2014) Reducing the negative vocal effects of superficial laryngeal dehydration with humidification. *Annals of Otology, Rhinology & Laryngology,* 123(7), pp. 475-481.

Lia, L., De, S. F., Mannocci, A., Di, V. N. and La, G. T. (2019) A cross sectional study on the vocal handicap index applied to a sample of teachers in nurseries and primary school. *Annali di igiene,* 31, pp. 230-235.

Liu, Z. W., Masterson, L. M., Srouji, I. A., Musonda, P. and Scott, D. G. I. (2012) Voice symptoms in patients with autoimmune disease: A cross-sectional epidemiological study. *Otolaryngology - Head and Neck Surgery,* 147(6), pp. 1108-1113.

Lubin, J. H., Alavanja, M. C. R., Caporaso, N., Brown, L. M., Brownson, R. C., Field, R. W., Garcia-Closas, M., Hartge, P., Hauptmann, M., Hayes, R. B., et al. (2007) Cigarette smoking and cancer risk: Modeling total exposure and intensity. *American Journal of Epidemiology,* 166(4), pp. 479-489.

Lyberg-Åhlander, V., Haake, M., Brännström, J., Schötz, S. and Sahlén, B. (2015a) Does the speaker's voice quality influence children's performance on a language comprehension test? *International Journal of Speech-Language Pathology,* 17(1), pp. 63-73.

Lyberg-Åhlander, V., Holm, L., Kastberg, T., Haake, M., Brännström, K. J., Sahlén, B., Cognitive, S., Lund, U., Logopedi, f. o. a., Logopedics, P., et al. (2015b) Are children with stronger cognitive capacity more or less disturbed by classroom noise and dysphonic teachers? *International Journal of Speech-Language Pathology,* 17(6), pp. 577-588.

Lyberg-Åhlander, V., Rydell, R., Fredlund, P., Magnusson, C. and Wilén, S. (2019) Prevalence of voice disorders in the general population, based on the Stockholm Public Health Cohort. *Journal of Voice,* 33(6), pp. 900-905.

Lynn, P. (2017) From standardised to targeted survey procedures for tackling non-response and attrition. *Survey Research Methods,* 11(1), pp. 93-103.

Ma, E. P. M. and Yiu, E. M. L. (2001) Voice activity and participation profile: Assessing the impact of voice disorders on daily activities. *Journal of Speech, Language, and Hearing Research,* 44(3), pp. 511-524.

Maasland, D. H. E., van den Brandt, P. A., Kremer, B., Goldbohm, R. A. and Schouten, L. J. (2014) Alcohol consumption, cigarette smoking and the risk of subtypes of head-neck cancer: results from the Netherlands Cohort Study. *BMC Cancer,* 14(1), pp. 187-187.

Mann, C. J. (2003) Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency Medicine Journal,* 20(1), pp. 54-60.

Manning, M., MacFarlane, A., Hickey, A. and Franklin, S. (2019) Perspectives of people with aphasia post-stroke towards personal recovery and living successfully: A systematic review and thematic synthesis. *PLOS ONE,* 14(3), p. e0214200.

Marmor, S., Horvath, K. J., Lim, K. O. and Misono, S. (2016) Voice problems and depression among adults in the U nited S tates. *The Laryngoscope,* 126(8), pp. 1859-1864.

Martin, S. and Darnley, L. (2004) *The Teaching Voice.* 2nd edn. London: Whurr.

Martinez, C. C. and Cassol, M. (2015) Measurement of voice quality, anxiety and depression symptoms after speech therapy. *Journal of Voice,* 29(4), pp. 446-449.
Martins, R. H. G., Gonçalvez, T. M., Pessin, A. B. B. and Branco, A. (2014a) Aging voice: presbyphonia. *Aging clinical and experimental research,* 26(1), pp. 1-5.
Martins, R. H. G., Pereira, E. R. B. N., Hidalgo, C. B. and Tavares, E. L. M. (2014b) Voice disorders in teachers: A review. *Journal of Voice,* 28(6), pp. 716-724.
Mason, G. (1987) Coping with collinearity. *The Canadian Journal of Program Evaluation,* 2, pp. 87-93.
Mathieson, L. and Greene, M. C. L. (2001) *Greene and Mathieson's the voice and its disorders.* London: Whurr.
McAleavy, G. J., Adamson, G., Hazlett, D. E., Donegan, H. A. and Livesey, G. E. (2008) Modelling determinants of the vocal health of teachers in Northern Ireland: Implications for educational policy and practice. *Public Health,* 122(7), pp. 691-699.
McAuley, L., Pham, B., Tugwell, P. and Moher, D. (2000) Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *The Lancet,* 356(9237), pp. 1228-1231.
McKinley, N., McCain, R. S., Convie, L., Clarke, M., Dempster, M., Campbell, W. J. and Kirk, S. J. (2020) Resilience, burnout and coping mechanisms in UK doctors: a cross-sectional study. *BMJ Open,* 10(1).
McNutt, L.-A., Wu, C., Xue, X. and Hafner, J. P. (2003) Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American Journal of Epidemiology,* 157(10), pp. 940-943.
Mela, C. F. and Kopalle, P. K. (2002) The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations. *Applied Economics,* 34(6), pp. 667-677.
Mentally Healthy Schools (2020) *Supporting Staff Wellbeing.* Web page. Available at: https://www.mentallyhealthyschools.org.uk/whole-school-approach/supporting-staff-wellbeing/ [Accessed 8th July 2020].
Mette, L. A., Pulido Saldívar, A., Poullard, N. E., Torres, I. C., Seth, S. G., Pollock, B. H. and Tomlinson, G. E. (2016) Reaching high-risk underserved individuals for cancer genetic counseling by video-teleconferencing. *Journal of Community and Supportive Oncology,* 14(4), pp. 162-168.
Millqvist, E., Bende, M., Brynnel, M., Johansson, I., Kappel, S., Ohlsson, A.-C., Institutionen för kliniska vetenskaper, s. f. h. p. o. ö., Institutionen för medicin, a. f. s. o. f., Institutionen för kliniska, v., Göteborgs, u., et al. (2008) Voice change in seasonal allergic rhinitis. *Journal of Voice,* 22(4), pp. 512-515.
Miri, A. K., Barthelat, F. and Mongeau, L. (2012) Effects of dehydration on the viscoelastic properties of vocal folds in large deformations. *Journal of Voice,* 26(6), pp. 688-697.
Misono, S., Haut, C., Meredith, L., Frazier, P. A., Stockness, A., Michael, D. D., Butcher, L. and Harwood, E. M. (2019) Dysphonia, perceived control, and psychosocial distress: A qualitative study. *Journal of Voice,* 33(5), pp. 682-690.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. and Group, T.P. (2009) Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Medicine, 6(7), p.e100097.
Molina, M. L., Fialho, A. R., Amaral, P. L. d., Bach, S. d. L., Rocha, L. M. d. and Souza, L. D. d. M. (2017) Well-being and associated factors among elementary school teachers in southern Brazil. *Revista CEFAC,* 19(6), pp. 812-820.
Moreti, F., Zambon, F., Oliveira, G. and Behlau, M. (2014) Cross-Cultural Adaptation, Validation, and cutoff values of the Brazilian version of the Voice Symptom Scale—VoiSS. *Journal of Voice,* 28(4), pp. 458-468.
Morgan, D. L. (2014a) *Integrating qualitative and quantitative methods: a pragmatic approach.* Los Angeles: SAGE.
Morgan, D. L. (2014b) Pragmatism as a paradigm for social research. *Qualitative inquiry,* 20(8), pp. 1045-1053.

Moy, F. M., Hoe, V. C. W., Hairi, N. N., Chu, A. H. Y., Bulgiba, A. and Koh, D. (2015) Determinants and effects of voice disorders among secondary school teachers in Peninsular Malaysia using a validated Malay version of VHI-10. *PLOS ONE,* 10(11).

NASUWT (2020) *Mental Health.* Web page. Available at: https://www.nasuwt.org.uk/news/campaigns/mental-health.html [Accessed 9th November 2020].

NatCen Social Research, U. C. L., Department of Epidemiology and Public Health (2019). *Health Survey for England, 2015*. [data collection]. 2nd Edition. UK Data Service. SN: 8280. Available at: http://doi.org/10.5255/UKDA-SN-8280-2 [Accessed 23 March 2018]

NatCen Social Research, U. C. L. D. o. E. a. P. H. (2015). *Health Survey for England, 2013.* [data collection]. UK Data Service. SN: 7649. Available at: http://doi.org/10.5255/UKDA-SN-7649-1 [Accessed 23 March 2018].

National Education Union (2019) *Voice Care for Teachers.* Web page. Available at: https://neu.org.uk/advice/voice-care-teachers [Accessed 8th July 2020].

Nerriere, E., Vercambre, M. N., Gilbert, F. and Kovess-Masfety, V. (2009) Voice disorders and mental health in teachers: a cross-sectional nationwide study. *BMC Public Health,* 9(1), pp. 370-370.

Nichol, H., Morrison, M. D. and Rammage, L. A. (1993) Interdisciplinary approach to functional voice disorders: The psychiatrist's role. *Otolaryngology—Head and Neck Surgery,* 108(6), pp. 643-647.

Nunnally, J. C. and Bernstein, I. (1994) *Psychometric Theory (McGraw-Hill Series in Psychology).* McGraw-Hill New York.

Nusseck, M., Immerz, A., Spahn, C., Echternach, M. and Richter, B. (2019) Long-term effects of a voice training program for teachers on vocal and mental health. *Journal of Voice*. In Press

Nusseck, M., Richter, B., Spahn, C. and Echternach, M., (2018). Analysing the vocal behaviour of teachers during classroom teaching using a portable voice accumulator. *Logopedics Phoniatrics Vocology*, 43(1), pp.1-10.Ohno, T., Hirano, S. and Rousseau, B. (2009) Age-associated changes in the expression and deposition of vocal fold collagen and hyaluronan. *Annals of Otology, Rhinology & Laryngology,* 118(10), pp. 735-741.

Oligschlaeger, Y., Yadati, T., Houben, T., Condello Oliván, C. M. and Shiri-Sverdlov, R. (2019) Inflammatory bowel disease: a stressed “Gut/Feeling”. *Cells,* 8(7), p. 659.

Özbal Koç, E. A., Koç, B. and Erbek, S. (2014) Comparison of acoustic and stroboscopic findings and voice handicap index between allergic rhinitis patients and controls. *Balkan Medical Journal,* 31(4), pp. 340-344.

Pan, B., Woodside, A. G. and Meng, F. (2014) How contextual cues impact response and conversion rates of online surveys. *Journal of Travel Research,* 53(1), pp. 58-68.

Paulsen, F., Kimpel, M., Lockemann, U. and Tillmann, B. (2000) Effects of ageing on the insertion zones of the human vocal fold. *Journal of anatomy,* 196(1), pp. 41-54.

Pernambuco, L., Espelt, A., Góis, A. C. B. and de Lima, K. C. (2017) Voice disorders in older adults living in nursing homes: Prevalence and associated factors. *Journal of Voice,* 31(4), pp. 510.e515-510.e521.

Phadke, K. V., Abo-Hasseba, A., Švec, J. G. and Geneid, A. (2019) Influence of noise resulting from the location and conditions of classrooms and schools in Upper Egypt on teachers' voices. *Journal Of Voice,* 33(5), pp. 802.e801-802.e809.

Pierce, J. L., Tanner, K., Merrill, R. M., Miller, K. L., Ambati, B. K., Kendall, K. A. and Roy, N. (2015) Voice disorders in Sjögren's syndrome: Prevalence and related risk factors. *Laryngoscope,* 125(6), pp. 1385-1392.

Pinto, C. R., Almeida, N. R., Marques, T. S., Yamamura, L. L. L., Costa, L. A. and Souza-Machado, A. (2013) Local adverse effects associated with the use of inhaled corticosteroids in patients with moderate or severe asthma. *Jornal Brasileiro de Pneumologia,* 39(4), pp. 409-417.

Portela, A. S., Granqvist, S., Ternström, S. and Södersten, M. (2018) Vocal behavior in environmental noise: comparisons between work and leisure conditions in women with work-related voice disorders and matched controls. *Journal of Voice,* 32(1), pp. 126. e123-126. e138.

Porter, S. R. and Whitcomb, M. E. (2007) Mixed-mode contacts in web surveys: Paper is not necessarily better. *Public opinion quarterly,* 71(4), pp. 635-648.

Puglisi, G. E., Astolfi, A., Cantor Cutiva, L. C. and Carullo, A. (2017) Four-day-follow-up study on the voice monitoring of primary school teachers: Relationships with conversational task and classroom acoustics. *The Journal of the Acoustical Society of America,* 141(1), pp. 441-452.

Ramaswamy, S., Rastogi, R. and Shim, K. (2000) *Efficient algorithms for mining outliers from large data sets*. Proceedings of the 2000 ACM SIGMOD international conference on Management of data 2000 May 16, pp. 427-438.

Rasbash, J. (2009) *What are multilevel models and why should I use them?* Online resource. Available at: http://www.bristol.ac.uk/cmm/learning/multilevel-models/what-why.html [Accessed 14th September 2020].

Reed, J. P. and Sims, H. S. (2016) Comparative analysis of characteristics of voice use amidst clergy. *Journal of Voice,* 31(2), pp. 256.e257-256.e211.

Revelle, W. (1979) Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research,* 14(1), p. 57.

Roberts, L. M., Wilson, S., Roalfe, A. and Bridge, P. (2004) A randomised controlled trial to determine the effect on response of including a lottery incentive in health surveys [ISRCTN32203485]. *BMC Health Services Research,* 4(1), pp. 30-30.

Rocha, B. R. and Behlau, M. (2018) The influence of sleep disorders on voice quality. *Journal of Voice,* 32(6), pp. 771. e771-771. e713.

Rogerson, J. and Dodd, B. (2005) Is There an effect of dysphonic teachers' voices on children's processing of spoken language? *Journal of Voice,* 19(1), pp. 47-60.

Rooney, L. K., Bhopal, R. S., Halani, L., Levy, M. L., Partridge, M. R., Netuveli, G., Car, J., Griffiths, C., Atkinson, J. P., Lindsay, G., et al. (2011) Promoting recruitment of minority ethnic groups into research: qualitative study exploring the views of South Asian people with asthma. *Journal of public health,* 33(4), pp. 604-615.

Rosen, C. A., Lee, A. S., Osborne, J., Zullo, T. and Murry, T. (2004) Development and validation of the voice handicap index-10. *Laryngoscope,* 114(9 I), pp. 1549-1556.

Rossi-Barbosa, L. A. R., Barbosa, M. R., Morais, R. M., de Sousa, K. F., Silveira, M. F., Gama, A. C. C. and Caldeira, A. P. (2016) Self-reported acute and chronic voice disorders in teachers. *Journal of Voice,* 30(6), pp. 755-755.e733.

Rothman, K. J., Greenland, S. and Lash, T. L. (2008) *Modern epidemiology.* Third edn. Philadelphia, PA: Lippincott Williams & Wilkins.

Roy, N. (2005) Teachers With Voice Disorders. *ASHA Leader,* 10(5), p. 8.

Roy, N., Merrill, R. M., Pierce, J. and Sundar, K. M. (2019) Voice disorders in obstructive sleep apnea: Prevalence, risk factors, and the role of CPAP. *Annals of Otology, Rhinology & Laryngology,* 128(3), pp. 249-262.

Roy, N., Merrill, R. M., Thibeault, S., Parsa, R. A., Gray, S. D. and Smith, E. M. (2004) Prevalence of voice disorders in teachers and the general population. *Journal of Speech, Language, and Hearing Research,* 47(2), pp. 281-293.

Roy, N., Tanner, K., Merrill, R. M., Wright, C., Miller, K. L. and Kendall, K. A. (2016) Descriptive epidemiology of voice disorders in rheumatoid arthritis: Prevalence, risk factors, and quality of life burden. *Journal of Voice,* 30(1), pp. 74-87.

Roy, N., Weinrich, B., Gray, S. D., Tanner, K., Toledo, S. W., Dove, H., Corbin-Lewis, K. and Stemple, J. C. (2002) Voice amplification versus vocal hygiene instruction for teachers with voice disorders: a treatment outcomes study. *Journal of Speech, Language & Hearing Research,* 45(4), pp. 625-638.

Royston, P., Moons, K. G. M., Altman, D. G. and Vergouwe, Y. (2009) Prognosis and prognostic research: Developing a prognostic model. *BMJ: British Medical Journal,* 338(7707), pp. 1373-1377.

Ruane, J. M. (2016) *Introducing social research methods: essentials for getting the edge.* Chichester, West Sussex, UK;Hoboken, NJ;: John Wiley & Sons Inc.

Ruel, E., Wagner III, W. E. and Gillespie, B. J. (2015) *The practice of survey research: Theory and applications.* Sage Publications.

Ruths, D. and Pfeffer, J. (2014) Social media for large studies of behavior. *Science,* 346(6213), pp. 1063-1063.

Saleh, A. and Bista, K. (2017) Examining Factors Impacting Online Survey Response Rates in Educational Research: Perceptions of Graduate Students. *Online Submission,* 13(2), pp. 63-74.

Saltürk, Z., Kumral, T. L., Bekiten, G., Atar, Y., Ataç, E., Aydoğdu, İ., Yıldırım, G., Kılıç, A. and Uyar, Y. (2016) Objective and subjective aspects of voice in pregnancy. *Journal of Voice,* 30(1), pp. 70-73.

Saltürk, Z., Özdemir, E., Kumral, T. L., Karabacakoğlu, Z., Kumral, E., Yildiz, H. E., Mersinlioğlu, G., Atar, Y., Berkiten, G. and Yildirim, G. (2017) Subjective and objective voice evaluation in Sjögren's syndrome. *Logopedics Phoniatrics Vocology,* 42(1), pp. 9-11.

Sampaio, M. C., dos Reis, E., Carvalho, F. M., Porto, L. A. and Araujo, T. M. (2012) Vocal Effort and Voice Handicap Among Teachers. *Journal of Voice,* 26(6), pp. 820.e815-820.e818.

Samuels, T. L., Handler, E., Syring, M. L., Pajewski, N. M., Blumin, J. H., Kerschner, J. E. and Johnston, N. (2008) Mucin gene expression in human laryngeal epithelia: Effect of laryngopharyngeal reflux. *Annals of Otology, Rhinology and Laryngology,* 117(9), pp. 688-695.

Sato, H. and Bradley, J. S. (2008) Evaluation of acoustical conditions for speech communication in working elementary school classrooms. *Journal of the Acoustical Society of America,* 123(4), pp. 2064-2077.

Sato, K. and Hirano, M. (1997) Age-related changes of elastic fibers in the superficial layer of the lamina propria of vocal folds. *Annals of Otology, Rhinology & Laryngology,* 106(1), pp. 44-48.

Sauermann, H. and Roach, M. (2013) Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features. *Research Policy,* 42(1), pp. 273-286.

Schiller, I. S., Morsomme, D., Kob, M. and Remacle, A. (2020) Noise and a speaker's impaired voice quality disrupt spoken language processing in school-aged children: evidence from performance and response time measures. *Journal of Speech, Language, and Hearing Research*, pp. 1-17.

Schmidt, A. F. and Finan, C. (2018) Linear regression and the normality assumption. *Journal of Clinical Epidemiology,* 98, pp. 146-151.

Schramm, S. H., Moebus, S., Lehmann, N., Galli, U., Obermann, M., Bock, E., Yoon, M.-S., Diener, H.-C. and Katsarava, Z. (2015) The association between stress and headache: A longitudinal population-based study. *Cephalalgia,* 35(10), pp. 853-863.

Scientific Advisory Committee of the Medical Outcomes Trust (2002) Assessing health status and quality-of life instruments: Attributes and review criteria. *Quality of Life Research,* 11(3), pp. 193-205.

Scott, S., Robinson, K., Wilson, J. A. and Mackenzie, K. (1997) Patient-reported problems associated with dysphonia. *Clinical Otolaryngology and Allied Sciences,* 22(1), pp. 37-40.

Seifert, E. (2005) Stress and distress in non-organic voice disorder. *Swiss Medical Weekly,* 135(2728).

Seifpanahi, S., Izadi, F., Jamshidi, A. A., Torabinezhad, F., Sarrafzadeh, J., Sobhani-Rad, D. and Ganjuie, K. (2016) Prevalence of voice disorders and associated risk factors in teachers and nonteachers in Iran. *Journal of Voice,* 30(4), pp. 506.e519-506.e523.

Sheikh, A., Halani, L., Bhopal, R., Netuveli, G., Partridge, M. R., Car, J., Griffiths, C. and Levy, M. (2009) Facilitating the recruitment of minority ethnic people into research: qualitative case study of South Asians and asthma. *PLoS Medicine,* 6(10).

Sheldon, H., Graham, C., Pothecary, N. and Rasul, F. (2007) Increasing response rates amongst black and minority ethnic and seldom heard groups. *Europe: Picker Institute*.

Sills, A. and Desai, P. (1996) Qualitative research amongst ethnic minority communities in Britain. *Journal of the Market Research Society,* 38(3), pp. 247-266.

Silvey, S. (1969) Multicollinearity and imprecise estimation. *Journal of the Royal Statistical Society: Series B (Methodological),* 31(3), pp. 539-552.

Simmons-Mackie, N. and Lynch, K. E. (2013) Qualitative research in aphasia: A review of the literature. *Aphasiology,* 27(11), pp. 1281-1301.

Sinclair, M., O'Toole, J., Malawaraarachchi, M. and Leder, K. (2012) Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. *BMC Medical Research Methodology,* 12(1), p. 132.

Siupsinskiene, N., Razbadauskas, A. and Dubosas, L. (2011) Psychological distress in patients with benign voice disorders. *Folia Phoniatrica et Logopaedica,* 63(6), pp. 281-288.

Sliwinska-Kowalska, M., Niebudek-Bogusz, E., Fiszer, M., Los-Spychalska, T., Kotylo, P., Sznurowska-Przygocka, B. and Modrzewska, M. (2006) The prevalence and risk factors for occupational voice disorders in teachers. *Folia Phoniatrica et Logopaedica,* 58(2), pp. 85-101.

Sloan, L., Morgan, J., Burnap, P. and Williams, M. (2015) Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PLOS ONE,* 10(3), p. e0115545.

Smith, E., Kirchner, H. L., Taylor, M., Hoffman, H. and Lemke, J. H. (1998a) Voice problems among teachers: Differences by gender and teaching characteristics. *Journal of Voice,* 12(3), pp. 328-334.

Smith, E., Lemke, J., Taylor, M., Kirchner, H. L. and Hoffman, H. (1998b) Frequency of voice problems among teachers and other occupations. *Journal of Voice - Official Journal of the Voice Foundation,* 12(4), pp. 480-480.

Smith, E., Verdolini, K., Gray, S., Nichols, S., Lemke, J., Barkmeier, J., Dove, H. and Hoffman, H. (1996) Effect of voice disorders on quality of life. *Journal of Medical Speech-Language Pathology,* 4(4), pp. 223-244.

Smith, G. (2008) Does gender influence online survey participation?: A record-linkage analysis of university faculty online survey response behavior. *ERIC Document Reproduction Service No. ED 501717*.

Smith, G. (2018) Step away from stepwise. *Journal of Big Data,* 5(1), p. 32.

Smith, S. L., Maxfield, L. and Hunter, E. J. (2019) Sensitivity analysis of muscle mechanics-based voice simulator to determine gender-specific speech characteristics. *Biomechanics and modeling in mechanobiology,* 18(2), pp. 453-462.

Snijders, T. A. B. (2005) Fixed and Random Effects. In: B.S. Everitt and D.C. Howell, ed. *Encyclopedia of statistics in behavioral science*, *2*. New Jersey: Wiley, pp. 664-665.

Spantideas, N., Drosou, E., Bougea, A. and Assimakopoulos, D. (2017) Inhaled corticosteroids and voice problems. What is new? *Journal of Voice,* 31(3), pp. 384. e381-384. e387.

Speyer, R., Speyer, I. and Heijnen, M. A. M. (2008) Prevalence and relative risk of dysphonia in rheumatoid arthritis. *Journal of Voice,* 22(2), pp. 232-237.

Stachler, R. J., Francis, D. O., Schwartz, S. R., Damask, C. C., Digoy, G. P., Krouse, H. J., McCoy, S. J., Ouellette, D. R., Patel, R. R., Reavis, C. C. W., et al. (2018) Clinical Practice Guideline: Hoarseness (Dysphonia) (Update). *Otolaryngology--head and neck surgery,* 158(1_suppl), p. S1.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M. and Carpenter, J. R. (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal,* 338, p. b2393.

Sterrett, D., Malato, D., Benz, J., Tompson, T. and English, N. (2017) Assessing changes in coverage bias of web surveys in the United States. *Public opinion quarterly,* 81(S1), pp. 338-356.

Stevens, D., Parkes, J. and Chan, S.-K. (2018) *ISC Census and Annual Report*. Kent. Online publication. Available at: https://www.isc.co.uk/media/4890/isc_census_2018_report.pdf [Accessed 12th July 2019].

Streiner, D. L. (2003) Starting at the Beginning: An introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment,* 80(1), pp. 99-103.

Sudman, S. and Bradburn, N. M. (1973) Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association,* 68(344), pp. 805-815.

Sun, G.-W., Shook, T. L. and Kay, G. L. (1996) Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *Journal of clinical epidemiology,* 49(8), pp. 907-916.

Sundarrajan, A., Huber, J. E. and Sivasankar, M. P. (2017) Respiratory and laryngeal changes with vocal loading in younger and older individuals. *Journal of Speech, Language, and Hearing Research,* 60(9), pp. 2551-2556.

Tan, S. H. and Tan, S. B. (2010) The correct interpretation of confidence intervals. *Proceedings of Singapore Healthcare,* 19(3), pp. 276-278.

Tanner, K., Roy, N. and Merrill, R. M. (2007) The effects of three nebulized osmotic agents in the dry larynx. *Journal of Speech, Language, and Hearing Research,* 50(3), p. 635.

Tavaluc, R. and Tan-Geller, M. (2019) Reinke's edema. *Otolaryngologic Clinics of North America,* 52(4), pp. 627-635.

Teddlie, C. and Tashakkori, A. (2009) *Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences.* London;Los Angeles, [Calif.];: SAGE.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J. and Stewart-Brown, S. (2007) The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes,* 5(1), p. 63.

The Association of Teachers and Lecturers (2008) *Teachers' Survey.* Web page. Available at: https://www.atl.org.uk/latest/press-release/68-teachers-primary-schools-have-suffered-voice-problems-atl [Accessed 4th July 2019].

Thibeault, S. L., Merrill, R. M., Roy, N., Gray, S. D. and Smith, E. M. (2004) Occupational risk factors associated with voice disorders among teachers. *Annals of Epidemiology,* 14(10), pp. 786-792.

Titze, I. R. (1988) The physics of small-amplitude oscillation of the vocal folds. *The Journal of the Acoustical Society of America,* 83(4), pp. 1536-1552.

Trinite, B. (2017) Epidemiology of voice disorders in Latvian school teachers. *Journal of Voice,* 31(4), p. 508.

Trudel-Fitzgerald, C., Millstein, R. A., von Hippel, C., Howe, C. J., Tomasso, L. P., Wagner, G. R. and VanderWeele, T. J. (2019) Psychological well-being as part of the public health debate? Insight into dimensions, interventions, and policy. *BMC Public Health,* 19(1), pp. 1-11.

Turley, R., Cohen, S. M., Becker, A. and Ebert Jr, C. S. (2011) Role of rhinitis in laryngitis: Another dimension of the unified airway. *Annals of Otology, Rhinology and Laryngology,* 120(8), pp. 505-510.

Ubillos, S., Centeno, J., Ibanez, J. and Iraurgi, I. (2015) Protective and risk factors associated with voice strain among teachers in Castile and Leon, Spain: Recommendations for voice training. *Journal of Voice,* 29(2), pp. 261.e261-261.e212.

University of London, Centre for Longitudinal Studies (2016) *1970 British Cohort Study: Forty-Two-Year Follow-Up, 2012* [data collection]. *2nd Edition.* UK Data Service. SN: 7473. Available at: http://doi.org/10.5255/UKDA-SN-7473-2. [Accessed 3rd March 2020]

University of London, Institute of Education, Centre for Longitudinal Studies (2015) *National Child Development Study Response and Outcomes Dataset, 1958-2013* [data collection]. 6th Edition. UK Data Service. SN: 5560. Available at: http://doi.org/10.5255/UKDA-SN-5560-4 [Accessed 3rd March 2020]

Van Houtte, E., Claeys, S., Wuyts, F. and Van Lierde, K. (2011) The impact of voice disorders among teachers: Vocal complaints, treatment-seeking behavior, knowledge of vocal care, and voice-related absenteeism. *Journal of Voice,* 25(5), pp. 570-575.

van Houtte, E., Claeys, S., Wuyts, F. and van Lierde, K. (2012) Voice disorders in teachers: Occupational risk factors and psycho-emotional factors. *Logopedics Phoniatrics Vocology,* 37(3), pp. 107-116.

Van Teijlingen, E. and Hundley, V. (2010) The importance of pilot studies. *Social research update,* 35(4), pp. 49-59.

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., Poole, C., Schlesselman, J. J., Egger, M. and Initiative, S. (2014) Strengthening the reporting of observational studies in epidemiology (STROBE): Explanation and elaboration. *International Journal of Surgery,* 12(12), pp. 1500-1524.

Verdolini-Marston, K., Titze, I. R. and Druker, D. G. (1990) Changes in phonation threshold pressure with induced conditions of hydration. *Journal of Voice,* 4(2), pp. 142-151.

Verdolini, K., Min, Y., Titze, I. R., Lemke, J., Brown, K., Mersbergen, M. v., Jiang, J. and Fisher, K. (2002) Biological mechanisms underlying voice changes due to dehydration. *Journal of Speech, Language, and Hearing Research,* 45(2), pp. 268-281.

Verdolini, K., Titze, I. R. and Fennell, A. (1994) Dependence of phonatory effort on hydration level. *Journal of speech and hearing research,* 37(5), pp. 1001-1007.

Ward, P. D., Thibeault, S. L. and Gray, S. D. (2002) Hyaluronic acid: its role in voice. *Journal of Voice,* 16(3), pp. 303-309.

Webb, A. L., Carding, P. N., Deary, I. J., MacKenzie, K., Steen, I. N. and Wilson, J. A. (2007) Optimising outcome assessment of voice interventions, I: Reliability and validity of three self-reported scales. *Journal of Laryngology and Otology,* 121(8), pp. 763-767.

Webb, P. and Bain, C. (2011) *Essential epidemiology: an introduction for students and health professionals.* 2nd edn. Cambridge: Cambridge University Press.

Weisstein, E. W. (2020) *Bonferroni Correction.* From MathWorld--A Wolfram Web Resource. Available at: https://mathworld.wolfram.com/BonferroniCorrection.html [Accessed 17th January 2020].

Whitling, S., Rydell, R. and Åhlander, V. L. (2015) Design of a clinical vocal loading test with long-time measurement of voice. *Journal of Voice,* 29(2), pp. 261. e213-261. e227.

Willems-Bloemer, L. H., Vreeburg, G. C. M. and Brummer, R. J. M. (2000) Treatment of reflux-related and non-reflux-related dysphonia with profound gastric acid inhibition. *Folia Phoniatrica et Logopaedica,* 52(6), pp. 289-294.

Williams, R. (2015) *Interaction effects between continuous variables (optional)*. Online publication. Available at: https://www3.nd.edu/~rwilliam/stats2/l55.pdf [Accessed 9th December 2020].

Willinger, U., Völkl-Kernstock, S. and Aschauer, H. N. (2005) Marked depression and anxiety in patients with functional dysphonia. *Psychiatry Research,* 134(1), pp. 85-91.

Wilson, J., Deary, I., Millar, A. and Mackenzie, K. (2002) The quality of life impact of dysphonia. *Clinical Otolaryngology & Allied Sciences,* 27(3), pp. 179-182.

Wilson, J. A., Webb, A., Carding, P. N., Steen, I. N., MacKenzie, K. and Deary, I. J. (2004) The Voice Symptom Scale (VoiSS) and the Vocal Handicap Index (VHI): a comparison of structure and content. *Clinical Otolaryngology,* 29(2), pp. 169-174.

Winefield, H. R., Gill, T. K., Taylor, A. W. and Pilkington, R. M. (2012) Psychological well-being and psychological distress: is it necessary to measure both? *Psychology of Well-Being: Theory, Research and Practice,* 2(1), p. 3.

WMA (2001) World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization,* 79(4), p. 373.

World Health Organisation (1980) *International classification of impairments, disabilities, and handicaps.* Geneva: World Health Organisation.

World Health Organisation (2001) *International Classification of Functioning, Disability and Health.* Geneva: World Health Organisation.

World Health Organisation (2013) *Investigating in mental health: evidence for action*. Geneva: World Health Organization. Online publication. Available at: https://www.who.int/mental_health/publications/financing/investing_in_mh_2013/en/ [Accessed 8th July 2019].

Wyss, A., Hashibe, M., Chuang, S.-C., Lee, Y.-C. A., Zhang, Z.-F., Yu, G.-P., Winn, D. M., Wei, Q., Sturgis, E. M., Talamini, R., et al. (2013) Cigarette, cigar, and pipe smoking and the risk of head and neck cancers: Pooled analysis in the international head and neck cancer epidemiology consortium. *American Journal of Epidemiology,* 178(5), pp. 679-690.

Yang, K., Tu, J. and Chen, T. (2019) Homoscedasticity: an overlooked critical assumption for linear regression. *General psychiatry,* 32(5).

Ylitalo, R., Li, W., Baugh, A. and Thibeault, S. (2004) Effect of acid and pepsin on gene expression in laryngeal fibroblasts. *Annals of Otology, Rhinology and Laryngology,* 113(11), pp. 866-871.

Zeileis, A., 2006. Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, 16(9), pp. 1-16.

Zellmer, S., Schmidt-Heck, W., Godoy, P., Weng, H., Meyer, C., Lehmann, T., Sparna, T., Schormann, W., Hammad, S. and Kreutz, C. (2010) Transcription factors ETF, E2F, and SP-1 are involved in cytokine-independent proliferation of murine hepatocytes. *Hepatology,* 52(6), pp. 2127-2136.

Appendices

Appendix 1 – Participant information

Participant Information

Research is being undertaken at Birmingham City University investigating the risk factors associated with voice problems in primary and secondary school teachers in England. We would like to invite you to participate in this research study. Before you decide to take part, it is important you understand why the research is being undertaken and what it will involve. Please take time to read the following information and decide whether you would like to take part. If you are not clear about anything or would like more information about the project, please contact Emily Sharp. Contact details are at the bottom of thepage.

What is the purpose of the study?

Current research suggests that teachers are more likely to report voice problems than the general population. The aim of this study is to investigate the risk factors, i.e. things such as health, psychological, lifestyle and behavioural factors that may lead to school teachers getting voice problems. Research has been carried out with teachers in other countries, but this research has not been of a high standard and therefore it is important that further research is undertaken. This is the first study of its kind to investigate the risk factors associated with voice problems for teachers in England. It is a nationwide study where all teachers in England will have the opportunity to take part. We would like all teachers, regardless of whether or not they have had a voice problem to take part in the research. This is because although it is also important to know why teachers get voice problems it is also important to know why some teachers do not. The study is due to be completed by October 2018.

Why have I been chosen?

You have been chosen because you are a teacher working in primary or secondary education in England. We have written to all schools and asked them to forward this invitation to every member of their teaching team.

Do I have to take part in the study?

No, your participation in this project is entirely voluntary. You are not obliged to take part. If you do agree to participate you have the right to withdraw from the study during the data collection period (2nd November 2017 - 28th December 2017) without having to give a reason. If you have completed the questionnaire but decide within this time that you do not want your information to be used your records will be destroyed.

What will I have to do if I agree to take part?

You will be asked to complete a web-based questionnaire which we estimate will take you approximately 15-25 minutes. If you cannot finish the questionnaire in one go you can save your answers and complete it later by clicking on the 'Finish Later' button which appears at the bottom of each page of the survey.

What are the possible benefits of taking part?

There are no immediate benefits to you personally taking part in the study. However, there will be benefits to teachers in general. The study will provide important information on why teachers are at risk of getting voice problems. This information will be useful to healthcare professionals in treating those with voice problems and also in helping to prevent voice problems from occurring. Results will be shared with health professionals and the teaching profession in order to inform their practice.

Are there any possible disadvantages in taking part?

Participating in the research is not anticipated to cause you any disadvantages or harm. However, although we think the risk is small some of you may be concerned about your vocal health or other health conditions after completing the questionnaire. If you do have any concerns about your health after completing the questionnaire, please contact your GP to discuss this further.

Will my taking part in this project be kept confidential?

The storage of data will comply with the Data Protection Act (1998). Any data collected about you in the online questionnaire will be stored online in a database protected by passwords and Birmingham City University security processes. Your data will only be used for the purposes of this research project and will only be accessible to the research team and will be kept strictly confidential. This means that your data will not be shared with anyone outside the research team. We will therefore not share with any schools or employers (or anyone else) any of your data, or whether you completed the questionnaire or not.

We plan to publish results from this research in summary form, but you and your school will not be identified in any reports or publications. We will also not ask you for your name or address in this study.

How long will my data be stored for?

10 years.

What will happen to the results of the research project?

We plan to publish results from this research in summary form, but you and your school will not be identified in any reports or publications. If you wish to be given a copy of any reports resulting from the research, please contact doctoral researcher Emily Sharp (for contact details see below).

Who is responsible for reviewing the study?

Dr Christel de Bruijn, Dr Wouter Jansen and Dr Salim Khan are responsible for reviewing the study.

The project has received ethical approval from the Health, Education and Life Sciences Faculty Ethics Committee of Birmingham City University.

Who can I contact for further information about the study?

Emily Sharp (Doctoral Researcher), Birmingham City University, City South Campus, Bevan 219, Faculty of Health, Education and Life Sciences, Westbourne Road, Edgbaston, B15 3TN.

Tel: 0121 331 7117

Email: Emily.sharp@bcu.ac.uk

What if I have a problem?

If you have a complaint or concern about any aspect of the project, please contact Dr Barbara Howard Hunt (Ethics) or Dr Christel de Bruijn (Director of Studies)

Dr Barbara Howard Hunt (Insurance Lead Ethics), Faculty of Health, Education and Life Sciences, Room 457 Seacole Building, Birmingham City University, Westbourne Road, Edgbaston, Birmingham B15 3TN

Tel: 0121 331 7162

Email: Barbara.HowardHunt@bcu.ac.uk

Dr Christel de Bruijn (Director of Studies), Department of Speech and Language Therapy and Rehabilitation, Faculty of Health, Education and Life Sciences, Birmingham City University, Westbourne Road, Edgbaston, Birmingham B15 3TN Tel: 0121 202 4218
Email: Christel.DeBruijn@bcu.ac.uk

Appendix 2 - Consent to Participate

The aim of this study is to investigate the risk factors that may lead to primary and secondary teachers getting voice problems. In order to do this, we are asking teachers in England who have a permanent contract with a school to complete an online survey.

We regret that for practical reasons we are unable to include Supply Teachers, Peripatetic Music Teachers or those that teach at Further Education level (Further Education is defined as study after A-levels that is not part of an undergraduate degree, and technical and applied qualifications for 16-19 year olds), or Nursery level (ages 0-4 Pre Reception) in this survey.

All aspects of this research are voluntary. You are not required to take part and can withdraw at any time while the data is being collected (2nd November 2017 - 28th December 2017)

If you would like to participate in this research, please read the following statements. If you agree with them please indicate that you consent to participate in the survey.

- I confirm that Ihave read the Participant Information for the above study.
- I have had the opportunity to consider the information, ask questions and had them answered satisfactorily.
- I understand that my participation is entirely voluntary and that I am free to withdraw my consent during (2nd November 2017- 28th December 2017) without giving a reason.
- I understand that all personal information will remain confidential and individuals or schools will not be identifiable by anyone else other than the research team. The names of individuals or schools will not appear in any published information.

I consent to participate in this survey □ *Required*

- ○ yes
- ○ no

Appendix 3 – Questionnaire

Introduction

As part of your job as a teacher, do you teach at all at Further Education level, at Nursery level or as a Supply Teacher or Peripatetic Music Teacher? Further Education is defined as study after A-levels that is not part of an undergraduate degree, and technical and applied qualifications for 16-19 year olds. Nursery level is defined as ages 0-4 (Pre Reception). *Required*

- ○ yes
- ○ no

PART 1 School Information

Please provide details of each school that you currently teach at. It is not important which school you list first.

It is important for the analysis of our results that we obtain the name of the school(s) that you teach at. We will never share any information from this survey with anyone from your school. Schools will not be named in any articles we may publish and schools and teachers will not be identifiable. We will also not tell your school whether or not you participated in this survey.

Name of school

Address

Town or City

Postcode

On average how many hours of teaching do you do each week at this school? Please note that we are interested in the actual number of teaching hours, not the total number of hours you are employed at this school.

Is this school a mainstream school (including fee paying and non-fee paying schools), special school or pupil referral unit/alternative provision? (Please select)

- Mainstream
- Special School
- Pupil Referral Unit/ Alternative Provision
- Don't know

Which of the following describes your school best?

- Community School
- Academy
- Free School
- Voluntary Aided
- Voluntary Controlled
- Foundation
- Private school (fee-paying school)
- Don't know

Is this school academically selective?

- yes
- no

Is this the only school you teach at?

- yes
- no

(If answers ‘no’, participant fills in the same information regarding their second school)

PART 2 Voice

This section of the questionnaire will ask you questions about your voice and the effects of your voice on your life.

Please answer all the following questions as best you can. It is important that we obtain an answer for each question. Even if there is no perfect response, please select the one that applies to you best. **When answering these questions think about how your voice has been in all areas of your life over the last 30 days.** *Required*

Please don't select more than 1 answer(s) per row.

Please select at least 30 answer(s).

	Never	Occasionally	Some of the time	Most of the time	Always
Do you have difficulty attracting attention?	☐	☐	☐	☐	☐
Do you have problems singing?	☐	☐	☐	☐	☐
Is your throat sore?	☐	☐	☐	☐	☐
Is your voice hoarse?	☐	☐	☐	☐	☐

When talking in company do people fail to hear you?	☐	☐	☐	☐	☐
Do you lose your voice?	☐	☐	☐	☐	☐
Do you cough or clear your throat?	☐	☐	☐	☐	☐
Do you have a weak voice?	☐	☐	☐	☐	☐
Do you have problems talking on the telephone?	☐	☐	☐	☐	☐
Do you feel miserable or depressed because of your voice?	☐	☐	☐	☐	☐
Does it feel as if there is something stuck in your throat?	☐	☐	☐	☐	☐
Do you have swollen glands?	☐	☐	☐	☐	☐
Are you embarrassed by your voice?	☐	☐	☐	☐	☐
Do you find the effort of speaking tiring?	☐	☐	☐	☐	☐
Does your voice make you feel stressed and nervous?	☐	☐	☐	☐	☐

Do you have difficulty competing against background noise?	☐	☐	☐	☐	☐
Are you unable to shout or raise your voice?	☐	☐	☐	☐	☐
Does your voice put a strain on your family and friends?	☐	☐	☐	☐	☐
Do you have a lot of phlegm in your throat?	☐	☐	☐	☐	☐
Does the sound of your voice vary throughout the day?	☐	☐	☐	☐	☐
Do people seem irritated by your voice?	☐	☐	☐	☐	☐
Do you have a blocked nose?	☐	☐	☐	☐	☐
Do people ask what is wrong with your voice?	☐	☐	☐	☐	☐
Does your voice sound creaky and dry?	☐	☐	☐	☐	☐
Do you feel you have to strain to produce voice?	☐	☐	☐	☐	☐
How often do you get throat infections?	☐	☐	☐	☐	☐

Does your voice ‘give out’ in the middle of speaking?	☐	☐	☐	☐	☐
Does your voice make you feel incompetent?	☐	☐	☐	☐	☐
Are you ashamed of your voice?	☐	☐	☐	☐	☐
Do you feel lonely because of your voice?	☐	☐	☐	☐	☐

PART 3 Teaching information

The next part of the questionnaire will ask you questions about your teaching practice. If you can't remember the exact details, please give your best guess.

Excluding prolonged periods of absence such as time spent on maternity/paternity leave, or sabbatical, how many years have you worked as a primary or secondary level teacher (do not include time spent as a student teacher)?

In the last 30 days which key stage have you spent the majority of your time teaching?

- ○ Early years
- ○ Key Stage 1
- ○ Key Stage 2
- ○ Key Stage 3
- ○ Key Stage 4
- ○ Key Stage 5

In the last 30 days which of the following subjects have you spent the majority of your time teaching? (please select one). If you are a Primary School or Early Years Classroom Teacher please select this option instead of choosing a subject.

Thinking back to **a typical week in the last 30 days** how many hours did you spend teaching this subject?

If applicable, **in the last 30 days** which of the following subjects did you spend the next most amount of time teaching? (please select one)

Please estimate the typical class size (number of pupils) you taught over the last 30 days.

In a typical week over the last 30 days what was the maximum number of hours you taught in a day?

In a typical week over the last 30 days what was the maximum number of hours you taught in a day **without a break?** A break is defined as half an hour or more without having to teach or supervise pupils, for example lunch time or non contact lessons.

In the last 30 days have you

Please don't select more than 1 answer(s) per row.

	Never	Sometimes	Most of the time	Always
had to talk louder than your normal speaking voice, in order to speak over noisy pupils?	☐	☐	☐	☐

taught in classrooms with a lot of background noise, e.g. from building or road works, traffic, classroom equipment or heating systems?	☐	☐	☐	☐
taught in classrooms that had a good acoustic, i.e. were comfortable to speak in?	☐	☐	☐	☐
taught in classrooms with air conditioning?	☐	☐	☐	☐

Please estimate the **number of hours in a typical week** that you use a microphone or other form of amplification while teaching.

Have you ever received information on how to look after your voice as part of the curriculum during your teacher training or as part of your professional development since you started working as a teacher?

- ○ Yes as part of my teacher training
- ○ Yes as part of my professional development
- ○ Yes as part of my teacher training and professional development
- ○ no

The following questions relate to activities that you take part in when not at work.

Please estimate **the number of occasions over the last 30 days** that you shouted, substantially raised your voice or strained your voice, for example at music festivals, sport events, rallies and public protests, bars and clubs with loud background noise.

Please estimate the **number of hours over the last 30 days** that you took part in activities that required you to use your voice over a sustained period of time, for example singing in a choir or a band, taking part in drama classes or amateur dramatics, public speaking.

How frequently over the last 30 days have you had to raise your voice at home? For example because you live or are frequently with someone who is hard of hearing, have children, have arguments with other people, have a dog, sing around the house

Please don't select more than 1 answer(s) per row.

	never	sometimes	often	most of the time	always
How often to you have to raise your voice at home?	☐	☐	☐	☐	☐

PART 4 Health

The next section will ask you questions about your health.

Have you ever been told by an Ear Nose and Throat (ENT) surgeon that you have a voice problem? (ENT surgeons are also known as laryngologists, otolaryngologists or otorhinolaryngologists)

- ○ yes
- ○ no
- ○ under investigation
- ○ don't know

The next few questions will ask you about heartburn and regurgitation. Please answer all questions. When you are answering the questions think about how how you have been over **the last 7 days.**

Please don't select more than 1 answer(s) per row.

	0 day	1 day	2-3 days	4-7 days

How often did you have a burning feeling behind your breastbone (heartburn)?	☐	☐	☐	☐
How often did you have stomach contents (liquid or food) moving upwards to your throat or mouth (regurgitation)?	☐	☐	☐	☐
How often did you have a pain in the centre of the upper stomach?	☐	☐	☐	☐
How often did you have nausea?	☐	☐	☐	☐
How often did you have difficulty getting a good night's sleep because of your heartburn and/or regurgitation?	☐	☐	☐	☐
How often did you take additional medication for your heartburn and/or regurgitation, other than what the doctor told you to take? (such as Rennie, Gaviscon, Nexium?)	☐	☐	☐	☐

In the last 30 days have you had a respiratory tract infection such as a cold, the flu, tonsillitis, sinusitis, laryngitis or bronchitis?

○ Yes confirmed by a doctor ○ I think so but have not seen a doctor ○ I don't think so

In the last 6 months have you had pneumonia?

○ yes ○ no ○ don't know

In the last 6 months have you had tuberculosis?

- yes
- no
- don't know

Do you have an allergy that causes a runny nose and/or breathing problems and/ or throat problems?

- Yes confirmed by a doctor
- I think I do but have not seen a doctor
- I don't think so

Has a doctor ever told you that you have asthma?

- yes
- no

In the last 30 days has there been an increase in your use of asthma medication or have you had to seek help from a health professional for your asthma?

- yes
- no

Have you ever been told by a doctor that you have fibromyalgia?

- yes
- no
- under investigation
- don't know

Have you ever been told by a doctor that you have rheumatoid arthritis?

- yes
- no
- under investigation
- don't know

Have you ever been told by a doctor that you have Sjören's syndrome?

- yes
- no
- under investigation
- don't know

Have you ever been told by a doctor that you have Myasthenia Gravis?

- yes
- no
- under investigation
- don't know

Have you ever been told by a doctor that you have Parkinson’s disease?

- yes
- no
- under investigation
- don't know

Have you ever been told by a doctor that you have Multiple Sclerosis?

- yes
- no
- under investigation
- don't know

Have you ever been told by a doctor that you have Cystic Fibrosis?

- yes
- no
- under investigation
- don't know

Have you ever had an injury or disease that causes you postural problems with your upper body, i.e. standing upright or sitting up straight?

- yes
- no
- don't know

Have you ever had an injury that has caused you damage in your throat or neck?

- ○ yes
- ○ no
- ○ don't know

Have you ever had a chest injury that causes you problems with your breathing?

- ○ yes
- ○ no
- ○ don't know

Have you ever been told by a doctor that you had the following types of cancer?

Please don't select more than 1 answer(s) per row.

	yes	no	under investigation
larynx/windpipe (also called trachea)	☐	☐	☐
mouth/tongue/lip	☐	☐	☐
throat/pharynx	☐	☐	☐
lung	☐	☐	☐

Have you ever received radiotherapy for the following types of cancer?

Please don't select more than 1 answer(s) per row.

	yes	no	not applicable
larynx/windpipe (also called trachea)	☐	☐	☐
mouth/tongue/lip	☐	☐	☐
throat/pharynx	☐	☐	☐
lung	☐	☐	☐

Have you ever had a thyroidectomy, i.e. surgery to remove all or part of your thyroid?

- ○ yes
- ○ no

Have you ever had endotracheal intubation, i.e. had a breathing tube inserted during surgery or intensive care?

- not applicable
- yes
- no
- don't know

Do you have any difficulty with your hearing?

- yes
- no

Which of the following best describes your menstrual status?

- Not applicable
- I haven't reached menopause yet
- I think I have reached menopause but this has not been confirmed by a doctor I have reached menopause and it has been confirmed by a doctor
- I am postmenopausal

Are you in the 3rd trimester of pregnancy?

- yes
- no
- not applicable

Have you been told by a doctor that you have polycystic ovary syndrome?

- yes
- no
- under investigation

Have you undergone surgery where both of your ovaries have been removed?

- yes
- no
- not applicable

Are you currently taking hormone replacement therapy to treat menopausal symptoms?

- yes
- no
- not applicable

How much do you weigh without your shoes? You can answer in pounds and stones **or** kilograms. Please only fill in one or the other.

Stones

and pounds

or kilograms

How tall are you without your shoes? You can answer in feet and inches **or** metres. Please only fill in one or the other.

feet and inches

or metres

PART 5 Wellbeing

Below are some statements about feelings and thoughts. Please mark the box that best describes your experience of each of these statements over the **last 2 weeks.**

Please don't select more than 1 answer(s) per row.

	None of the time	Rarely	Some of the time	Often	All of the time
I've been feeling optimistic about the future	☐	☐	☐	☐	☐
I've been feeling useful	☐	☐	☐	☐	☐
I've been feeling relaxed	☐	☐	☐	☐	☐
I've been feeling interested in other people	☐	☐	☐	☐	☐

I've had energy to spare	☐	☐	☐	☐	☐
I've been dealing with problems well	☐	☐	☐	☐	☐
I've been thinking clearly	☐	☐	☐	☐	☐
I've been feeling good about myself	☐	☐	☐	☐	☐
I've been feeling close to other people	☐	☐	☐	☐	☐
I've been feeling confident	☐	☐	☐	☐	☐
I've been able to make up my own mind about things	☐	☐	☐	☐	☐
I've been feeling loved	☐	☐	☐	☐	☐
I've been interested in new things	☐	☐	☐	☐	☐
I've been feeling cheerful	☐	☐	☐	☐	☐

PART 6 Lifestyle

Now some questions about smoking and drinking

Which of these statements apply to you?

- ○ Ismoke cigarettes every day
- ○ I now smoke cigarettes occasionally but not every day I
- ○ used to smoke cigarettes but don't at all now
- ○ I've never smoked cigarettes

On days that you smoke how many cigarettes do you usually have?

Do you currently use e cigarettes (vaping)?

- yes
- no

How often do you currently use an electronic cigarette?

- Daily
- Less than daily, but at least once a week Less
- than weekly, but at least once a month Less
- than monthly

How often do you have an alcoholic drink of any kind?

- On most days
- 2 to 3 days a week
- once a week
- 2 to 3 times a month
- once a month
- Less often or only on special occasions Never
- nowadays

In the last 7 days, how many units of alcohol did you drink? A unit is half a pint of beer, a small glass of wine or a single measure of spirits and liqueur.

PART 7 Demographic Information

This section of the questionnaire asks for demographic information. Please answer all questions as best you can.

What was your age on your last birthday?

What sex were you assigned at birth as stated on your birth certificate?

- male
- female
- prefer not to say

Are you currently undergoing or have you undergone gender transitioning? (by gender transitioning we mean the adopting of a different gender than the one assigned at birth. This does not necessarily include medical treatment such as surgery or drugs)

- yes
- no
- prefer not to say

Have you undergone voice surgery as part of gender transitioning?

- not applicable
- yes

- no
- prefer not to say

What is your ethnic group? Choose one option that best describes your ethnic group or background.

- White British or Irish
- Any other White background
- Asian Indian, Pakistani or Bangladeshi Asian
- Chinese
- Any other Asian background
- Black African
- Black Caribbean
- Any other Black background
- Mixed White and Asian
- Mixed White and Black
- Any other Mixed background
- Any other ethnic group

Prefer not to say

PART 8 Your household

There has been research to suggest a link between health outcomes and wealth. In order to obtain this information we would like to ask you questions about your household and your income.

What type of household do you live in? A household is one person living alone, or a group of people (not necessarily related) living at the same address who share cooking facilities and living room or sitting room or dining areas.

- couple with no children living with you
- couple with dependent children only living with you
- couple with dependent and independent children living with you couple
- with independent children only living with you
- lone parent with dependent children only living with you
- lone parent with dependent and independent children living with you lone
- parent with independent children only living with you
- two or more families living in the same household
- lone person sharing with other lone persons
- one male one

Please provide details of the monthly net income of your household (that is income after tax, National Insurance pension contributions and union subscriptions) including any overtime, bonuses, tips, commission and tax refund. Please provide this information to the nearest pound. If you are not able to provide this information please write 'don't know' in the box.

Please provide the monthly income from other sources your household receives, including state benefits, income from pensions, income from private sources such as rent from lodgers and income from savings and investments. Please provide this information to the nearest pound.

How much council tax do you pay every month? Please answer to the nearest pound.

How many adults aged 18 and over (other than yourself) live in your household?

How many children aged 14 and above live with you?

How many children aged 0-13 live with you?

Many thanks for completing the questionnaire

Many thanks for completing this questionnaire. Please retain the reference number that you have been given. Please quote this reference number if you need to contact us regarding the study.

Please email: emily.sharp@bcu.ac.uk

We hope we have not given you any cause for concern about your health or wellbeing, but in case you do have any concerns, please contact your GP.

Appendix 4 – Ethical Approval Letter

Faculty of Health, Education and Life Science Research Office

Faculty of Health, Education and Life Sciences

Birmingham City University

Westbourne Road

Birmingham
B15 3TN

HELS_Ethics@bcu.ac.uk

15/08/2017
Ms Emily Sharp
Westbourne Road
Edgbaston
Birmingham
B15 3TN
United Kingdom

Dear Ms Emily Sharp

Re: Risk Factors For Voice Problems in Primary and Secondary School Teachers in England - Sharp /Aug /2017 /RLRA /1312

Thank you for your application and documentation regarding the above study. I am happy to take Chair's Action and approve the study, following the adjustments you have made, meaning that you may now begin your research.

The Committee's opinion is based on the information supplied in your application. If you wish to make any substantial changes to the research please contact the Committee and provide details of what you propose to alter. A substantial change is one that is likely to affect the

- Safety and well-being of the participants
- Scientific value of the study
- Conduct or management of the study

The Committee should also be notified of any serious adverse effects arising as a result of this research. The Committee is required to keep a favourable opinion under review in the light of progress reports.

I wish you every success with your study.

Yours sincerely,

Mr. Stuart Mitchell

On behalf of the Faculty Academic Ethics Committee

Appendix 5 – Introductory email to Head Teachers

Introductory email.

Subject Header: Nationwide Study - Voice Problems in Teachers

Dear Head Teacher,

Could you please forward this email to all teachers in your school?

Dear Teacher,

Current research suggests that school teachers are more likely to report voice problems than the general population. I am a researcher at Birmingham City University and part of a team that is undertaking research to investigate which risk factors, (e.g. health, psychological, environmental and behavioural) may lead to voice problems in primary and secondary school teachers.

This is a nationwide study where all school teachers in England will have the opportunity to take part. The information you provide will allow the research team to understand why some teachers get voice problems while others do not.

To ensure that our findings are representative of all teachers in England, it is important that we have as many completions of the online survey as possible. We thus invite you to complete our online survey which should take 15-25 minutes. Your responses will be completely confidential. Details about the study and whether you are eligible to take part can be found at the beginning of the survey.

To start the survey please click on the link below. We invite you to complete the survey as soon as possible, although the survey will be available to complete until midnight 31st December 2018.

https://bcu.onlinesurveys.ac.uk/nationwide-study-voice-problems-teachers

If you have any questions about the research or experience any technical difficulties please contact emily.sharp@bcu.ac.uk or telephone 0121 331 7117.

Thank you for your time in assisting with this important research.

Kind regards

Emily Sharp

Birmingham City University

Faculty of Health, Education and Life Sciences

Westbourne Road

Birmingham B15 3TN