ANNUAL REVIEW OF EDUCATION, COMMUNICATION, AND LANGUAGE SCIENCES

https://research.ncl.ac.uk/arecls/

SPECIAL ISSUE (2021)

Special issue volume contents

Editorial Welcome

On behalf of the Annual Review of Education, Communication and Language Sciences team, we would like to welcome you to this special issue. Our theme is 'Intervention and Research Practice', which highlights the current innovative studies which Newcastle University's postgraduates and staff have completed. We were very fortunate to receive five papers which exemplify this theme, and present them to you as follows:

Joanna Baker is a newly qualified Speech and Language Therapist. She is currently completing a MSc in Clinical Linguistics and Evidence Based Practice (Research) at the School of Education Communication and Language Sciences in Newcastle University after being awarded a 1+3 NINE DTP studentship. Her research interests include phonological awareness, speech sound disorders and cleft palate. Stephanie van Eeden is currently the Lead Specialist Speech and Language Therapist for the Newcastle Cleft Lip and Palate Service and is undertaking a PhD researching the linguistic and auditory skills of children born with cleft lip and palate. Dr Helen Stringer is currently Head of Speech and Language Sciences at Newcastle University. She is an academic speech and language therapist with broad research interests that include speech sound disorders and behaviour change theories. They compared the speech and language outcomes of two groups, children with isolated cleft palate and children with cleft palate and Robin Sequence. The data were retrospective case-note reviews obtained for children at 18 months, 3 years old and 5 years old. Their results are presented here, alongside further research recommendations for longitudinal investigation into language outcomes for these groups.

Jack Charnley is a doctoral candidate in the School of Education, Communication and Language Sciences at Newcastle University, attached to the Water Security and Sustainable Development Hub. As a qualified primary school teacher and projects manager for an international development charity, Jack is particularly interested in how young children engage with Water, Sanitation and Hygiene (WASH) environments as conscious, responsible, resilient and informed citizens. His article outlines his investigation into the principal factors affecting children's WASH behaviours and health outcomes in East New Delhi primary schools and how these serve to mitigate or exacerbate WASH-related inequalities, driving or impeding progress towards Sustainable Development Goal 6.

Sally J McDonald read Education with English at Queens' College, Cambridge, and the completed a Masters degree in Educational Leadership at the University of Warwick. She is now studying for an EdD at Newcastle University, researching how trainee teachers' placements affect their training experiences and consequently their retention in teaching. Sally is interested in all aspects of educational inequality and social justice, particularly how to make education more equitable by improving teachers' training and professional development to support staff wellbeing, retention and school improvement, especially in disadvantaged areas. She has also

researched the educational experiences of school-aged mothers. She works as an English teacher, tutor and examiner. She is also a primary school governor and has been involved in a range of initiatives to improve access to Higher Education. Her progress report concerns her current EdD project, which uses interviews and the participant database from Teach First to examine how Teach First's placement strategy is a) affecting experiences of participants in the programme, b) impacting the participants and schools partnered with Teach First (from the perspective of school staff), and c) affects the relationship with retention on the programme. She then provides preliminary findings, and discusses these, in addition to outlining changes to her method, and the potential implications of this due to the covid-19 pandemic.

Hanh Pho completed her PhD in Educational and Applied Linguistics at School of Education, Communication and Language Sciences at Newcastle University. She was born in Vietnam, a Southeast Asian country, and so understands deeply the struggle of an international student in a foreign context, and how this experience can permanently alter the worldview of a student. Her research interests include language, culture and identity; intercultural communication; cross-cultural transition, adjustment and adaptation; and internationalisation in higher education. Her article presents research which examines the different forms of social grouping that occurs with international students in taught postgraduate degree settings, and explores why these formations may occur from their perspectives. The work is completed via semi-structured interviewing at the beginning and end of the academic year.

Yiyin Wang is a second year PhD student studying Applied Linguistics in the School of Education, Communication and Language Sciences at Newcastle University. In China, he used to work as a lecturer at Zhejiang International Studies University in Hangzhou. Before joining the PhD program, he taught International English Language Testing System (IELTS) speaking courses for more than ten years. His current research focuses on comparing two video rating modes of the IELTS speaking test, which is supervised by Professor Paul Seedhouse. His article is a literature review on the IELTS test, providing an overview of its four key elements and exploring the validity of the test. He then provides recommendations for research which takes each into account the test developers, administrators and instructors' special interest into concern.

We would also like to take this time to thank the student and staff editors who worked hard on helping draw this issue together. Our thanks go to Nadia Ahmed, Badryah Almesfer, Nada Bin Ghali, Chang Liu, Lee Robinson, Yao Wang and Paul Seedhouse. We would also like to give special thanks to Simin Ren, who was the previous senior editor and has been invaluable at guiding Josie to be able to produce this issue. Finally, a special thanks to the contributing authors of this issue, who have conducted some incredible work and have given us the opportunity to share it here.

We hope that you enjoy this special issue, and please let us know if you have any feedback for the journal. Our call for volume 18(2) will be 6th August 2021 and we

encourage staff and students to submit if their areas of interest include Education, Communication or Language Sciences.

Josie Tulip (Senior Student Editor) and Peter Sercombe (Editor in Chief)

June, 2021

Early speech and language outcomes in non-syndromic cleft palate with and without Robin Sequence: A matched case study

Joanna Baker, Stephanie van Eeden and Helen Stringer

ABSTRACT: Robin Sequence (RS) is a rare condition leading to orofacial anomalies, including cleft palate. Evidence suggests children with RS have poor long-term speech and language outcomes, even in comparison to those with isolated cleft palate. This study compared speech and language outcomes of children with isolated cleft palate (ICP) to children with cleft palate and RS (CPRS). A retrospective case-note review was completed of 74 matched children (37 ICP; 37 CPRS). Speech and language assessment at 18-months, three and five years provided data for palatal function and articulation. At age five years, children with CPRS had significantly more frequent and severe articulation errors than children with ICP. Palatal function across the two groups was not significantly different. Expressive and receptive language at three years did not differ. Results support previous studies highlighting the severity and frequency of articulation errors associated with CPRS. Further longitudinal investigation into language outcomes is recommended.

Key words: *Robin Sequence, cleft palate, speech, articulation, velopharyngeal insufficiency*

Contact: J.Baker@newcastle.ac.uk

Introduction

Robin Sequence (RS) is a congenital condition which occurs from 1 in 5600 to 1 in 30,000 children (Paes, van Nunen, Bassert & Breugem, 2015; Caouette-Laberge, Bayet & Larocque, 1994). It is heterogeneous, with an often-differing developmental prognosis across individuals (Thouvenin et al., 2013). In 2016, Breugem et al. produced a clinical consensus report regarding the characteristics of RS; micrognathia (small lower jaw) was determined the primary and initiating characteristic, with glossoptosis (retraction of the tongue) and airway obstruction cited as mandatory diagnostic characteristics. Although not included in the diagnostic criteria, cleft palate is most commonly Stickler's Syndrome[1] and 22q11 deletion syndrome[2] (Izumi et al, 2012).

Children with cleft palate and RS (CPRS) and children with cleft palate only, termed isolated cleft palate (ICP) in this paper, often require support to aid their speech and language

[1] Stickler syndrome is a disorder characterised by a distinctive facial appearance, hearing loss, joint problems and eye abnormalities.

[2] 22q11 deletion syndrome, also termed DiGeorge syndrome, is caused by the deletion of a segment of chromosome 22. Symptoms of 22q11 deletion syndrome often include congenital heart problems, developmental delay, learning difficulties and cleft palate or velopharyngeal incompetence.

development and commonly follow the same treatment pathway. Current evidence suggests that children with CPRS have increased speech and language difficulties when compared to children with ICP, but studies are sparse and often have samples including syndromic and non-syndromic RS (Filip et al., 2015; Thouvenin et al., 2013). Speech difficulties related to cleft palate include ongoing issues with palatal function resulting in velopharyngeal insufficiency (VPI). This leads to hypernasal resonance and nasal airflow errors accompanying consonant production. Compensatory and obligatory articulation errors can also develop. These are often referred to as Cleft Speech Characteristics (CSCs). These include anterior errors (e.g. lateral and palatal placement); posterior errors (e.g. backing to velar or uvular); non-oral errors (e.g. active nasal fricatives, glottal or pharyngeal placement); and passive errors (e.g. weak or nasalised errors related to ongoing structural difficulties with the velopharyngeal mechanism) (John, Sell, Harding-Bell, Sweeney & Williams, 2006).

Velopharyngeal insufficiency

Velopharyngeal Insufficiency (VPI) often occurs after repair of the cleft palate (Goudy, Ingraham & Canady, 2011) and some studies suggest that children with CPRS experience increased VPI. In a longitudinal prospective study, Thouvenin et al. (2013) assessed the speech of 39 children with RS (27 with isolated RS and 12 with RS and Stickler's syndrome) at 15 months and at three and six years. They analysed the long-term developmental outcomes of the children; at six-years 69% (N=25) had persistent hypernasal speech. In parity with this study, Filip et al. (2015) investigated a total of 93 non-syndromic individuals with CPRS over a 33-year period, collecting data retrospectively and comparing it with a large control group. After cleft palate repair 30 of 87 (34.5%) participants with non-syndromic and syndromic RS developed normal speech while 33.3% of 93 non-syndromic patients had surgery for VPI. This is a significantly higher rate when compared to the control group of individuals with ICP, of which only 19.4% received further surgery. Hardwicke, Richards, Cafferky, Underwood, ter Horst and Slator (2016) assessed articulation and nasality outcomes of 24 patients with RS at five years of age. These outcomes were compared with an ICP control group. Hardwicke et al. (2016) report that the RS group had significantly higher rates of secondary surgery ($p=0.017$) and significantly poorer nasality scores ($p=0.031$) than the ICP group. In another retrospective study of 130 children with isolated or syndromic CPRS, Morice et al. (2018) found similar rates of VPI (30.5%) in their isolated CPRS group at a mean age of 3.4 years.

Findings are mixed regarding VPI outcomes and not all studies report significant differences when comparing children with CPRS to children with ICP. Goudy, Ingraham and Canady (2011) state that children with RS do not have a higher rate of VPI after cleft palate repair and that they are no more likely to require additional surgery than children with ICP. However, in 2021,

Schwaiger et al. produced the largest study reporting speech outcomes using an established standardised assessment tool within a follow-up time frame, comparing 51 patients with CPRS to an ICP cohort at age five years. They found that children with CPRS were more likely to need further surgery to correct VPI before age five years (p=0.016). In a systematic review by Wan et al. (2015), only six papers were found comparing the speech outcomes of children with RS compared to ICP. Four of the papers which measured VPI reported no significant outcomes. Meanwhile, Stranksy et al. (2013) found a significant difference in VPI between children with RS and those with ICP at age eight (p=0.04).

Articulation

Current research has compared articulation outcomes of children with CPRS, although often in small samples. In Wan et al.'s (2015) systematic review, only three papers studied articulation outcomes (Stranksy et al, 2013; Lehman, Fishman & Neiman, 1995; Khosla, Mabry & Castiglione, 2008). Their descriptions of error types were limited, and no significant group differences were found. In a matched comparison study of 24 children with RS and 24 with ICP, Hardwicke et al. (2016) found that the RS group had poorer articulation outcomes; they had significantly worse cleft speech characteristics (CSCs) (p=0.023) and produced more backed to velar errors, glottal substitutions and active nasal fricatives than the ICP group. The number of each CSC type found were not provided within the paper and only 24 participants were included in the study. In the larger Schwaiger et al. (2021) paper children with RS presented with a significantly higher rate of CSCs (p=0.001). Overall, the small number of papers, and the small sample sizes within the studies limits the application of results to clinical practice. Therefore, results are tentative, but do suggest that articulation outcomes at the age of five years after cleft palate repair were poorer in children with RS than for children with ICP.

Language difficulties

Children with cleft palate often perform poorly on assessments of language compared to their peers. A recent scoping review concluded that there is evidence of early language delay in children born with cleft lip and palate (van Eeden & Stringer, 2020). However, research investigating the language difficulties of children with RS/CPRS is scarce, and the existing literature has conflicting outcomes. One study by Smith et al. (2014) explored the outcomes related to sleep disturbance caused by obstructive sleep apnoea in children with CPRS at three years of age. Significant differences were found for both receptive (p=0.01) and expressive language (p=0.001) using the Bayley Scales of Infant and Toddler Development – III[3] (Bayley, 2006)

[3] The Bayley Scales of Toddler Development is an assessment that is used to measure five areas of child development including cognition, motor skills, language, social-emotional abilities and adaptive behaviour.

when compared to children with ICP. Alencar et al. (2017) investigated 62 children with isolated RS and cleft palate, exploring whether airway management by positioning or nasopharyngeal airway tube affected rates of hypoxia which may lead to cognitive difficulties. They found that between 18.4% and 20.8% of children with CPRS were at risk of language difficulties. The Thouvenin et al. (2013) study discussed above, also analysed the long-term developmental outcomes of the children. At 15 months of age, the language scores for the children were below average, and at three years of age, 18 children (46%) still had language delay and their mean score on vocabulary subtest was below one standard deviation. However, no follow-up scores at six-years were reported.

Purpose of this study

There are limited longitudinal data regarding speech and language development in children with CPRS making it difficult to understand the speech and language difficulties of this population. Not enough is known about children with CPRS and how they differ from children with ICP. Research outcomes lack consensus with largely mixed results regarding speech and language outcomes. Further research is necessary to better understand the treatment and management required by children with CPRS for successful speech and language development. Comparing children with CPRS to those with ICP in further investigations will increase our understanding of speech and language development for children with RS. The present study adds to the current evidence base by comparing the early speech and language outcomes of children with CPRS to a matched group of children with ICP, with the aim of informing clinical practice, enabling clinicians to identify which children are at risk of later speech and language difficulties.

Research Questions

This study aimed to answer the following questions:

Compared to non-syndromic children with isolated cleft palate, do children with non-syndromic Robin Sequence have:

1. Worse speech outcomes at five years in terms of velopharyngeal function and CSCs present in their articulation?

2. Worse early speech and language outcomes at 18 months and three years?

Secondary research questions explored correlations of speech outcomes with severity of cleft; presence of fistula post-surgery; history of hearing difficulties; expressive language at three years; receptive language at three years.

Ethical approval

Ethical approval was granted through the Newcastle University Education, Communication and Language Sciences ethics committee (REF. ECLS_317). NHS approvals in line with Caldicott principles were granted by Newcastle upon Tyne Hospitals

NHS Foundation Trust (REF. NUTH/7141).

Method

Design

This was a retrospective, longitudinal, matched case study. Participants were taken from the caseload of a regional cleft lip and palate service in the United Kingdom. They were matched by socioeconomic status, gender and age at final speech assessment.

Table 1. Sample comparisons across the matched groups

Group	Mean (SD) age at final speech assessment (months)	Socioeconomic status (Mean IMD (SD))	Gender ratio (M:F)
Isolated Cleft Palate	64 (2.24)	4.03 (2.87)	17:20
Cleft Palate + Robin Sequence	65 (2.11)	4.43, (2.26)	16:21
Chi-square analysis	X^2 (10, N=74) = 7.71, p=.657	X^2 (8, N=74) = 9.15, p=.330	X^2 (1, N=74) = 0.55, p=.815; Fisher's Exact (2-sided) p=.100

Participants

All participants were born between 01/01/2004 and 31/12/2013. All had surgery carried out by a single surgeon. Chi-squared analysis confirmed group comparability (see Table 1). There was a total of 74 participants with 37 in each group.

Data collection

All data was stored electronically on a spreadsheet before being entered into IBM SPSS Statistics (IBM, 2016).

Primary outcome measures

Data from a five year audit assessment regarding velopharyngeal function and CSCs was collected following the Cleft Audit Protocol for Speech–Augmented (CAPS-A) protocol (John et al., 2006). The CAPS-A is used throughout the United Kingdom to audit speech outcomes at five-years and is a robust measure of velopharyngeal function and articulation outcomes derived through intense training of speech and language therapists (SLTs), with ongoing calibration and a consensus

listening process (Sell et al. 2009). Data included measures of resonance, nasal airflow and the presence of anterior, posterior, non-oral or passive CSCs.

Secondary outcome measures

Data identifying rates of secondary surgery was collected. Early speech and language data were also investigated. Data on resonance and presences of CSCs at 18 months and three years were collected from the speech and language records along with data from language assessments collected at three years. This was taken from informal observation and spontaneous speech samples using methods described in the Derbyshire Language Scheme (Knowles & Masidlover, 1982). For example, understanding at one-word level was rated as a severe delay at three-years, whilst understanding of three or more information-carrying words in an instruction was considered as normal for their age. The equivalent was used for expressive language; an expressive level of only single words at age three was regarded as a severe delay.

Confounding variables

The following data were collected from medical records to investigate any associations with the primary outcome measures and confounding variables: severity of the cleft; age at primary palate repair, history of fistulae following surgery; hearing history.

Data analysis

IBM SPSS Statistics for Windows, Version 21.0was used for all data analysis. Data on resonance, nasal airflow and non-oral CSCs were combined using the VPC-SUM method described in Lohmander et al., 2017. This is calculated on a 3-point scale to describe velopharyngeal function (0 = competent, 1 = borderline, 2 = incompetent). Data on CSCs were collated into seven levels in order of severity – none, anterior, posterior, non-oral, anterior and posterior, anterior and non-oral, all three types. Frequency statistics were obtained for speech and language outcomes at 18 months, 3 years and 5 years, as well as secondary surgery, fistulae rates and hearing history. Group differences were analysed using a non-parametric Mann-Whitney U test, using a p-value of <0.05 as level of significance. Correlations, using Spearman's Rho (one-tailed), were made to investigate the associations of identified early risk factors with the primary outcomes; these were the secondary outcomes and confounders outlined above. Further inferential statistics using binary logistic regression were carried out to analyse the extent to which poor speech outcomes could predict to which group a participant belonged.

Results

Primary outcome measures

Speech results at five years were taken as the primary outcome measure, as reported by consensus among specialist SLTs working in the field of cleft palate and velopharyngeal dysfunction.

For velopharyngeal function, 76% of the CPRS group had competent VP function compared with 92% of the ICP group. This difference was not statistically significant (p=.055). However, a weak, significant negative correlation was seen between velopharyngeal function at five years and whether a child was in the CPRS or ICP group (r=-.225, p=.027), meaning that there was an association between a diagnosis of RS and the likelihood of ongoing VPI. Binary logistic regression analysis showed that poor velopharyngeal function at five years did not significantly predict whether a participant had a diagnosis or RS or not (B=.786, Exp(B)=2.195, p=.106). For articulation, only 32% of the CPRS group had good articulation with no evidence of CSCs at age five; this compared to 65% of the ICP group and was found to be highly significant (p=.002). A weak to moderate negative correlation was seen between articulation at five years and whether a child was in the CPRS or ICP group (r=-.367, p=.001), meaning that there was an association between a diagnosis of RS and significant articulation difficulties. Binary logistic regression showed that an increased level of articulation difficulty made it more likely a participant had a diagnosis of RS (B=.472, Exp(B)=1.602, p=.018).

Secondary outcome measures

For secondary surgery, a total of 8 children had had secondary surgery for speech by the time of their five year speech assessment. There were 6 in the RS group (16%) and 2 in the ICP group (5%). This did not represent a significant difference (Fishers Exact 2-sided, p=.261). For language scores, language measures were only available at 3-years. Language samples from which to gather speech development data at 18-months differed between the groups, with only 41% of the CPRS group with enough language at this age compared with 59% of the ICP group. For expressive language skills, 46% of the CPRS group were considered to be delayed, compared with 38% of the ICP group. This difference was not statistically significant (Fisher's Exact 2-sided, p=.621). For receptive language skills, 43% of the CPRS group were considered to have a delay, compared with 31% of the ICP group. Again, this was not statistically significant (Fisher's Exact 2-sided, p=.449). Although there were no significant differences seen between groups, there was a weak positive correlation with articulation at five years and both expressive language (r=.243, p=.024) and receptive language (r=.281, p=.011) at three years. This means there was an association between poorer language skills at age three years and articulation difficulties at five years.

Confounding variables

Regarding severity of the cleft, there was a significant difference across the groups in terms of the type of cleft palate; 97% of the CPRS group had a cleft of the hard and soft palate compared with only 54% of the ICP group (p<.001). For age at palate repair, there was a significant difference in the range of ages at which palates were repaired with the

majority of those in the CPRS group being repaired at around 12-months and those in the ICP group at 6-months (p=.006). With presence of a fistula after primary repair, there was a significant difference across groups with regard to fistulae with 32% of the CPRS group having a fistula after primary repair compared to 8% of the ICP group (p=.018). As for history of hearing problems, no significant group differences were seen here; 68% of the CPRS group had a history of hearing problems compared to 62% of the ICP group (p=.603). Weak to moderate significant correlations were seen between velopharyngeal function at five years and post-surgical fistulae (r=.329, p=.006) and severity of the cleft (r=.248, p=.033). Weak to moderate significant correlations were seen between articulation at five years and post-surgical fistulae (r=.388, p=.001), severity of the cleft (r=.365, p=.001) and hearing difficulties (r=.258, p=.033). The binary logistic regression showed that of the confounding variables, only the severity of the cleft palate was found to be significant when adjusting for the influence of both velopharyngeal function at five years (B=3.346, Exp(B)=28.379, p=.002) and articulation at five years (B=.3.173, Exp(B)=23.876, p=.004) on group membership.

Longitudinal data

Speech data was collected at 18-months, three years and five years in order to observe patterns of development between the two groups. At all ages the CPRS group had poorer outcomes.

Velopharyngeal function - At 18 months of age only 15 of the children in the CPRS group had a large enough spontaneous language sample to assess VP function, compared with 22 in the ICP group. All those that had samples to analyse were rated as having competent VP function using the VPC-SUM. At 3 years, 73% of the children in the CPRS group had competent VP function, compared to 94% in the ICP group. This was found to be a significant difference (p=.029). As reported above, by five years the gap had closed a little to 76% of the CPRS group having competent VP function compared with 92% of the ICP group and was no longer a significant difference (p=.055). See Figure 1. At 18 months of age 60% of the children in the CPRS group were not showing signs of developing CSCs in their speech; this compared with 100% in the ICP group. This was highly significant (p=.001). At 3 years, only 53% of the children in the CPRS group had good articulation with no CSCs, compared to 73% in the ICP group. By this age the group differences seen were non-significant (p=.053). But as reported above, by five years the rates of CSCs had increased in both groups and the gap had widened again with only 32% of the CPRS group having good articulation with no CSCs, compared with 65% of the ICP group (p=.002). See Figure 2 for further breakdown of the types of CSCs reported at the different ages

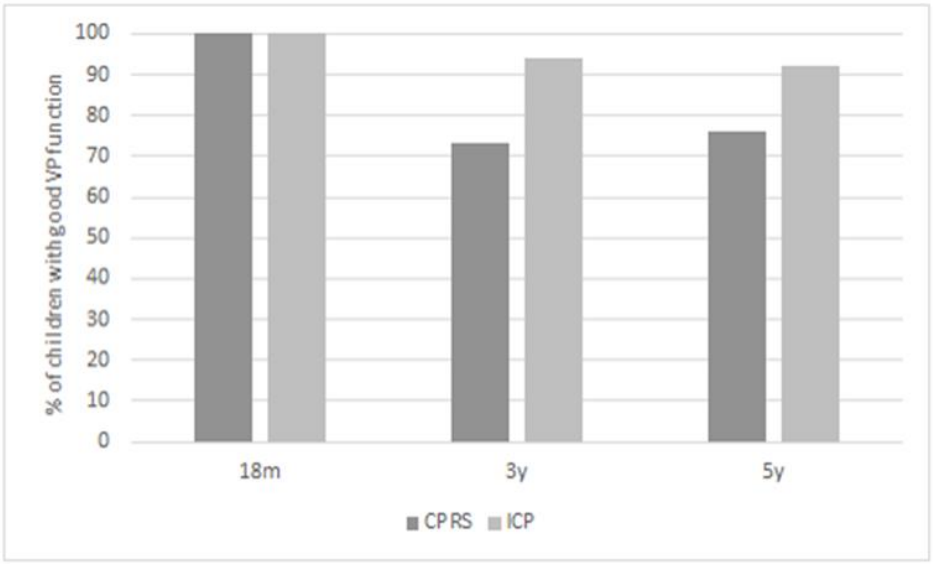


Figure 1. Longitudinal data regarding velopharyngeal function

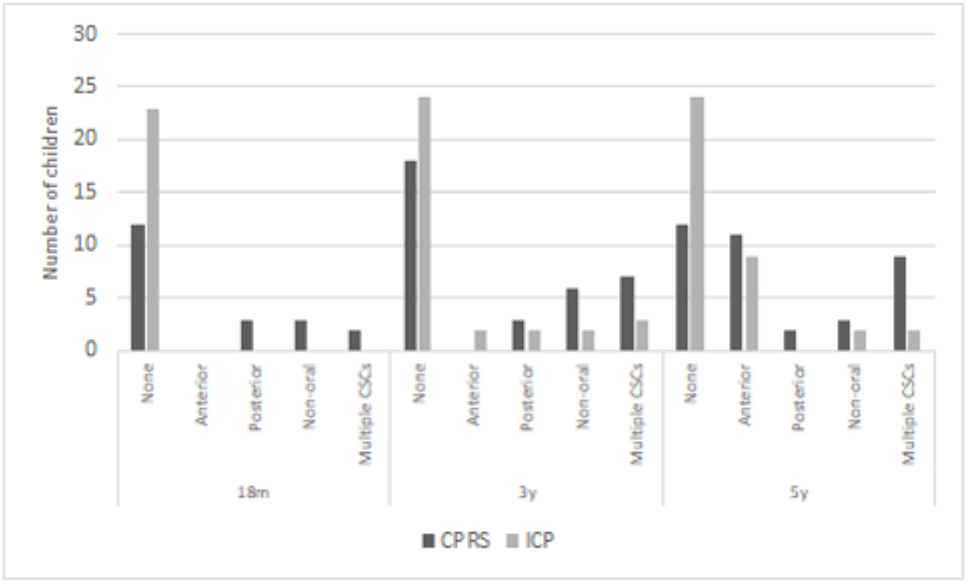


Figure 2. Longitudinal data for articulation and breakdown of CSCs

Across all the ages studied, the CPRS group had poorer outcomes, with the gap between this group and peers with ICP widening for articulation skills as they reached five years.

Discussion

In the present study, speech outcomes at the age of five years after cleft palate repair were poorer in children with CPRS than for children in the matched ICP comparison group. The

CPRS group performed significantly worse in terms of articulation. Comparison can be made between the present study and results published by Hardwicke et al. (2016) and Schwaiger et al. (2021). All three studies had comparable surgical protocols and used the same validated speech assessment tool (John et al., 2006). Consistent with the present study, both Hardwicke et al. (2016) and Schwaiger et al. (2021) found statistically significant differences in CSCs between CPRS and ICP cohorts, reporting that CPRS participants performed worse than their ICP counterparts.

Palatal function at age 5-years between the two groups was not significantly different in the present study, which is in harmony with recent reports (Schwaiger et al., 2021). Overall, more children with CPRS had VPI than the ICP group; 24% of children with CPRS presented with mild or severe VPI, which is slightly below recently reported VPI rates for non-syndromic CPRS by Morice et al. (2018) (30.5%), Stransky et al. (2013) (47%) and Hardwicke et al. (2016) (46%). Interestingly, prior studies have shown mixed results when comparing VPI outcomes in children with CPRS to children with ICP. Thouvenin et al. (2013)'s longitudinal study involved a similar sample size, to the present study yet included syndromic cases of CPRS. Nevertheless, they found contrasting outcomes. At the age of six years, children with CPRS had significantly increased VPI compared to children with ICP (Thouvenin et al, 2013). Several additional studies comparing CPRS with ICP support this conclusion (Goudy, Ingraham & Canady, 2011; Khosla, Mabry & Castiglione, 2008; Hardwicke et al., 2016; Morice et al., 2018). It is however unclear from these studies whether any secondary surgery for speech had taken place. In this study six children in the CPRS group had already had secondary surgery by the age of five which influenced the velopharyngeal function outcome.

Expressive and receptive language outcomes at three years did not differ significantly between the CPRS and ICP groups. These outcomes add to the discussion regarding language difficulties in RS, although results remain limited. Though not found in the present study, significant differences for both receptive and expressive language have been reported in recent years when comparing children with RS to those with ICP (Smith et al., 2014). In a longitudinal study, Thouvenin et al. (2013) found that 18 of 39 children with RS still had language delay at age three years. However, 12 of the 39 children within Thouvenin et al.'s (2013) study had RS associated with Stickler's syndrome. In the present study, children with CPRS and associated syndromes were excluded, which could account for the differences in results. Whilst a significant difference between the ICP and CPRS groups was not observed in this study, more children with CPRS did have language difficulties, at a rate comparable to that in the Thouvenin et al. study (46% at three years). Therefore, it could be postulated that children with CPRS have more language difficulties than children with ICP and that this was not detected statistically due to methodological

consequences. For example, there is potential that the language outcomes reported in the present study were impacted by the assessments used within the service. There is reason to believe that the language assessments used may not have been sensitive enough to detect a difference in difficulties between the two groups; decisions regarding language were based on spontaneous speech samples using a non-standardised measure which may not have captured subtle but important differences in the children's abilities. Moreover, it is important to note that language data for some participants was missing, meaning language results were not representative of the whole cohort. Whilst this notion regarding language results is not entirely without substance, it may be a result of researcher bias linked to preconceptions of previous literature and study expectations. Overall, language outcomes from this study are tentative and must be considered with caution. Further investigation and longitudinal follow-up are warranted to explore whether language results were impacted by the assessment used and the absence of data.

Strengths and Limitations

Comparison groups have been used previously in research. Studies have compared syndromic versus non-syndromic RS cohorts (Wagener 2003; de Buys 2008; Witt 1997). In addition, comparisons have been made between non-RS cohorts and RS cohorts both matched and unmatched (Stranksy, 2013; Goudy, Ingraham & Canady, 2011; Lehman, Fishma & Neima, 1995), not dissimilar to the present study. The sample size achieved within this study (37 in each group), although larger than existing studies (Hardwicke et al., 2016; Schwaiger et al., 2021), remains relatively small. Goudy et al. (2011) suggest that matching by age, sex, age at primary surgery and cleft anatomy forms a valuable comparison group; participants with CPRS were matched in terms of socio-economic status, gender, date of birth and age at 5-year speech audit, which strengthens the group comparison. Moreover, age at primary surgery was accounted for in the regression analysis and was not found to significantly contribute to the outcomes. However, the severity of the cleft was found to significantly influence the outcomes for both groups. It is not possible to tell from this study whether the diagnosis of RS contributes to poorer speech outcomes or whether the severity of the cleft has a greater influence. Future studies should ensure that groups are matched by severity of the cleft.

Prior findings suggests that there would be a much higher rate of syndromic CPRS patients if they were genetically screened (Izumi et al, 2012). Although none of the participants in the present study had a diagnosed syndrome, it is acknowledged that there may be participants with undiagnosed syndromes within the ICP and/or CPRS groups, which could have impacted study results. Contrasting outcomes have been found comparing syndromic versus non-syndromic RS patients. No statistically significant differences in speech outcomes

between syndromic and non-syndromic RS patients were observed in Schwaiger et al.'s 2021 case series whereas Patel et al. (2012) and more recently Logjes et al. (2021) report significantly worse speech outcomes in syndromic RS patients.

Most prior comparative studies have been retrospective and completed within single establishments, which is the case for the present study. The retrospective nature meant that recordkeeping could not be controlled, and treatment fidelity could not be promoted. Whether participants received intervention or support over the years monitored was not included, which could have impacted speech and language outcomes. As information was gathered from case files retrospectively, participants were excluded because their five year speech audit was missing or incomplete. This occurred in prior studies (Hardwicke et al., 2016) and could be deemed an element of selection bias.

Conclusion

This study highlights the severity of cleft speech characteristics in children with Robin Sequence. Children with CPRS experience worse articulation outcomes when compared to children with ICP, and likely require increased support in this area. In agreement with prior research, results demonstrate the vulnerability of this client group. Too little evidence exists regarding this rare congenital condition in terms of speech and especially language outcomes. There is a need for large scale multi-disciplinary studies to allow for the development of specific standards and processes to support children with RS.

References

Alencar, T. R. R., Marques, I. L., Bertucci, A., Prado-Oliveira, R. (2017). Neurological Development of Children with Isolated Robin Sequence Treated with Nasopharyngeal Intubation in Early Infancy. *The Cleft Palate-Craniofacial Journal, 54*(3), 256–261. https://doi.org/10.1597/14-228

Bayley, N. (2006). *Bayley scales of infant and toddler development*. San Antonio, TX: The Psychological Corporation.

Breugem, C. C., Evans, K. N., Poets, C. F., Suri, S., Picard, A., Filip, C., . . . Hong, P. (2016). Best Practices for the Diagnosis and Evaluation of Infants With Robin Sequence: A Clinical Consensus Report. *JAMA pediatrics, 170*(9), 894-902. doi:10.1001/jamapediatrics.2016.0796

Caouette-Laberge, L., Bayet, B. & Larocque, Y. (1994). The Pierre Robin sequence: review of 125 cases and evolution of treatment modalities. *Plast Reconstr Surg, 93*(5):934-42. PMID: 8134485.

de Buys Roessingh, A. S., Herzog, G., Cherpillod, J., Trichet-Zbinden, C., & Hohlfeld, J. (2008). Speech prognosis and need of pharyngeal flap for non syndromic vs syndromic Pierre Robin Sequence. *J Pediatr Surg, 43*(4), 668-674. doi:10.1016/j.jpedsurg.2007.09.050

Filip, C., Feragen, K. B., Lemvik, J. S, Lindberg, N., Andersson, E, Rashidi, M., … Høgevold, H. E. (2015). Multidisciplinary Aspects of 104 Patients with Pierre Robin Sequence. *The Cleft Palate-Craniofacial Journal*,

52(6), 732–742. https://doi.org/10.1597/14-161

Giudice, A., Barone, S., Belhous, K., Morice, A., Soupre, V., Bennardo, F., . . . Picard, A. (2018). Pierre Robin sequence: A comprehensive narrative review of the literature over time. *Journal of stomatology, oral and maxillofacial surgery, 119*(5), 419-428. doi:10.1016/j.jormas.2018.05.002

Goudy, Steven, Ingraham, Christopher, & Canady, John. (2011). The occurrence of velopharyngeal insufficiency in Pierre Robin Sequence patients. *International Journal of Pediatric Otorhinolaryngology, 75*(10), 1252–1254. https://doi.org/10.1016/j.ijporl.2011.06.024

Gustafsson, C., Vuola, P., Leikola, J., & Heliovaara, A. (2020). Pierre Robin Sequence: Incidence of Speech-Correcting Surgeries and Fistula Formation. *Cleft Palate Craniofac J, 57*(3), 344-351. doi:10.1177/1055665619874991

Haapanen, M. L., Laitinen, S., Paaso, M., & Ranta, R. (1996). Quality of speech correlated to craniofacial characteristics of cleft palate patients with the Pierre Robin sequence. *Folia Phoniatr Logop, 48*(5), 215-222. doi:10.1159/000266412

Holder-Espinasse, M., Abadie, V., Cormier-Daire, V., Beyler, C., Manach, Y., Munnich, A., . . . Amiel, J. (2001). Pierre Robin Sequence: A series of 117 consecutive cases. *The Journal of Pediatrics, 139*(4), 588-590. doi:10.1067/mpd.2001.117784

Humphries, L. S., & Taylor, J. A. (2021). Discussion: Robin Sequence: 5-Year Speech Outcomes-A Case-Control Study. *Plastic and Reconstructive Surgery, 147*(3), 687-688. doi:10.1097/PRS.0000000000007694

Izumi, K. M. D., Konczal, L. L. M. D., Mitchell, A. L. M. D., & Jones, M. C. M. D. (2012). Underlying Genetic Diagnosis of Pierre Robin Sequence: Retrospective Chart Review at Two Children's Hospitals and a Systematic Literature Review. *The Journal of Pediatrics, 160*(4), 645-650.e642. doi:10.1016/j.jpeds.2011.09.021

John A. K., Sell D., Harding-Bell A., Sweeney T., Williams A. (2006) CAPS-A: a validated and reliable measure for auditing cleft speech. *Cleft Palate–Craniofacial Journal; 43*: 272–288

Khosla, Rohit K, Mabry, Kelly, & Castiglione, Charles L. (2008). Clinical Outcomes of the Furlow Z-Plasty for Primary Cleft Palate Repair. *The Cleft Palate-Craniofacial Journal, 45*(5), 501–510. https://doi.org/10.1597/07-063.1

Knowles, W. and Masidlover, M. (1982) The Derbyshire Language Scheme.

LeBlanc, S. M., & Golding-Kushner, K. J. (1992). Effect of glossopexy on speech sound production in Robin sequence. *The Cleft Palate-craniofacial Journal, 29*(3), 239-245. doi: 10.1597/1545-1569_1992_029_0239_eogoss_2.3.co_2

Lehman, J. A., Fishman, J. R. A., Neiman, G. S. (1995). Treatment of Cleft Palate Associated with Robin Sequence: Appraisal of Risk Factors. *The Cleft Palate-craniofacial Journal, 32* (1), 25-29. Doi: https://doi.org/10.1597/1545-1569_1995_032_0025_tocpaw_2.3.co

2

Logjes, R. J. H., Upton, S., Mendelsohn, B. A., Badiee, R. K., Breugem, C. C., Hoffman, W. Y., & Pomerantz, J. H. (2021). Long-term Speech Outcomes of Cleft Palate Repair in Robin Sequence versus Isolated Cleft Palate. *Plastic and reconstructive surgery. Global open, 9*(1), e3351-e3351. doi:10.1097/GOX.0000000000003351

Lohmander, A., Hagberg, E., Persson, C., Willadsen, E., Lundeborg, I., Davies, J., . . . Nyberg, J. (2017). Validity of auditory perceptual assessment of velopharyngeal function and dysfunction - the VPC-Sum and the VPC-Rate. *Clinical linguistics & phonetics, 31*(7-9), 589-597. doi:10.1080/02699206.2017.1302510

Morice, A., Renault, F., Soupre, V., Chapuis, C., Trichet Zbinden, C., Kadlub, N., . . . Picard, A. (2018). Predictors of speech outcomes in children with Pierre Robin sequence. *Journal of Cranio-maxillofacial Surgery, 46*(3), 479-484. doi:10.1016/j.jcms.2017.12.004

Paes, Emma C, van Nunen, Daan P. F, Basart, Hanneke, Don Griot, J. Peter W, van Hagen, Johanna M, van der Horst, Chantal M. A. M, … Breugem, Corstiaan C. (2015). Birth prevalence of Robin sequence in the Netherlands from 2000-2010: a retrospective population-based study in a large Dutch cohort and review of the literature. *American Journal of Medical Genetics*, *167*(9), 1972–1982. https://doi.org/10.1002/ajmg.a.37150

Patel, K. B., Sullivan, S. R., Murthy, A. S., Marrinan, E., & Mulliken, J. B. (2012). Speech outcome after palatal repair in nonsyndromic versus syndromic Robin sequence. *Plastic and Reconstructive Surgery, 130*(4), 577e-584e. doi:10.1097/PRS.0b013e318262f2e4

Pinto, M. D. B., Pegoraro-Krook, M. I., Andrade, L. K. F., Correa, A. P. C., Rosa-Lugo, L. I., & Dutka, J. C. R. (2017). Intensive treatment of speech disorders in robin sequence: a case report. *Codas, 29*(5), e20160084. doi:10.1590/2317-1782/20172016084

Prado-Oliveira, R., Marques, I. L., Souza, L., Souza-Brosco, T. V., & Dutka Jde, C. (2015). Assessment of speech nasality in children with Robin Sequence. *CoDAS, 27*(1), 51-57. doi:10.1590/2317-1782/20152014055

Schwaiger, M., Cook, H., Jordan, Z., Edmondson, S. J., Mischak, I., Wallner, J., . . . Ferguson, L. (2021). Robin Sequence: 5-Year Speech Outcomes-A Case-Control Study. *Plastic and Reconstructive Surgery, 147*(3), 676-686. doi:10.1097/PRS.0000000000007641

Sell, D., John, A., Harding-Bell, A., Sweeney, T., Hegarty, F. & Freeman, J. (2009) Cleft Audit Protocol for Speech (CAPS-A): a comprehensive training package for speech analysis. *International Journal of Language & Communication Disorders, 44*:4, 529-548, DOI: 10.1080/13682820802196815

Smith, C. B., Walker, K., Badawi, N., Waters, K. A., & MacLean, J. E. (2014). Impact of sleep and breathing in infancy on outcomes at three years of age for children with cleft lip and/or palate. *Sleep, 37*(5), 919–925. https://doi.org/10.5665/sleep.3660

Thouvenin, B., Djadi-Prat, J., Chalouhi, C., Pierrot, S., Lyonnet, S., Couly, G. &

Abadie, V. (2013). Developmental outcome in Pierre Robin sequence: A longitudinal and prospective study of a consecutive series of severe phenotypes. *American Journal of Medical Genetics, 161*(2), 312–319. https://doi.org/10.1002/ajmg.a.35773

van Eeden, S. & Stringer, H. (2020). Linguistic and auditory processing skills in non-syndromic children with cleft palate: A scoping review. *Journal of Communication Disorders, 87*, 106029–106029. https://doi.org/10.1016/j.jcomdis.2020.106029

van Lieshout, M. J. S., Joosten, K. F. M., Hoeve, H. L. J., Mathijssen, I. M. J., Koudstaal, M. J., & Wolvius, E. B. (2014). Unravelling Robin sequence: considerations of diagnosis and treatment. *The Laryngoscope, 124*(5), E203-E209. doi:10.1002/lary.24437

Wagener, S., Rayatt, S., Tatman, A. J., Gornall, P., & Slator, R. (2003). Management of infants with Pierre Robin sequence. *The Cleft Palate-Craniofacial Journal, 40*(2), 180–185. https://doi.org/10.1597/1545-1569(2003)0402.0.CO;2

Wan, T., Chen, Y., & Wang, G. (2015). Do patients with isolated Pierre Robin Sequence have worse outcomes after cleft palate repair: A systematic review. *Journal of plastic, reconstructive & aesthetic surgery, 68*(8), 1095-1099. doi:10.1016/j.bjps.2015.04.015

Witt, P. D., Myckatyn, T., Marsh, J. L., Grames, L. M., & Dowton, S. B. (1997). Need for velopharyngeal management following palatoplasty: an outcome analysis of syndromic and nonsyndromic patients with Robin sequence. *Plastic and Reconstructive Surgery, 99*(6), 1522–1529. https://doi.org/10.1097/00006534-199705010-00008

Sustaining Water, Sanitation and Hygiene (WASH) behaviour change produced during schools-based interventions: Perspectives from East New Delhi primary schools

Jack Charnley

ABSTRACT: The United Nations' Sustainable Development Goal 6 (SDG 6) aims to eradicate damaging behaviours related to Water, Sanitation and Hygiene (WASH). Schools-based WASH interventions can contribute to this, rendering the school environment a safe space for children to practise healthy WASH behaviours. However, those which focus on infrastructural upgrades unaccompanied by a sufficient behavioural change element do not maximise their efficacy. One country which faces numerous challenges in achieving SDG 6 is India, and implicit within the SDG targets and existing evidence is the relationship between this, quality education (SDG 4) and long-term impacts on children's physical and cognitive development. It is important that behavioural change is implemented successfully to address such issues. This research considers how the benefits from schools-based WASH interventions in East Delhi primary schools can be maximised to sustain children's WASH behaviours post-implementation. Through grounded theory analysis of primary qualitative data (semi-structured interviewing and ethnographic observations including photography and a daily reflective diary), it considers how the Capability, Opportunity, Motivation Behaviour (COM-B) model can be applied to ensure children's sustained adoption of healthy WASH behaviours. The research shows that schools-based WASH interventions have the potential to provide the opportunity for this but identifies barriers to children's capability to adopt WASH behaviours. It also suggests that school staff lack the necessary motivation to support children in long-term positive WASH behaviour adoption. Teachers have an important role as WASH ambassadors but they themselves need support in acquiring the capability, opportunity and motivation to fulfil this.

Key words: *Behaviour change, New Delhi, primary schools, SDG 6, WASH*

Contact: j.charnley1@ncl.ac.uk

Introduction

Reaching the United Nations' Sustainable Development Goal[1] (SDG) 6 (Clean Water and Sanitation) is a particular challenge in India, which continues to have one of the highest rates of open defecation (defecating in the open rather than in a toilet or latrine) in the world (Ritchie and Roser, 2019). Over 7% of the population lives without basic water access (UNICEF,

[1] The SDGs are the United Nations' 2030 targets for global development and aim to address many factors which pose a threat to global development (United Nations General Assembly, 2015). There are 17 in total, covering themes such as 'No Poverty' (SDG 1), 'Zero Hunger' (SDG 2) and 'Good Health and Well-Being' (SDG 3).

2020). Limited access to sanitation provision in low-income communities further embeds the practice of open defecation, the elimination of which is central to SDG 6; one of its targets is to "achieve access to adequate and equitable sanitation and hygiene for all and end open defecation" by 2030 (United Nations General Assembly, 2015, p. 18). Damaging WASH-related behaviours such as open defecation and insufficient handwashing are exacerbated by a lack, non-use or misuse of sanitation facilities, running water and soap. They are directly linked to the faecal-oral transmission of dangerous respiratory and gastrointestinal diseases including environmental enteropathy, cholera and Escherichia coli which are the leading worldwide causes of infant mortality (Humphrey, 2009; Mara, 2017; Troeger et al, 2017; Khalil et al, 2018). More recently, COVID-19 can also be added to the list of deadly diseases which spread through inadequate access to WASH, particularly in low-income areas where overcrowding is an issue and water provision is intermittent (Brauer et al, 2020; Ray, 2020). As of June 2021, India has the second highest number of cases in the world and the third highest number of deaths (Johns Hopkins University, 2021).

Inadequate WASH services widen pre-existing inequalities as predominantly low-income children face WASH insecurity which has long-term impacts on their physical and cognitive development. Stunting, "the impaired growth and development that children experience from poor nutrition, repeated infection, and inadequate psychosocial stimulation" (World Health Organization, 2015, para. 1), is a particular problem in India where, as of 2017, over one-third of children under the age of five were affected (World Bank, 2020). Stunting is positively correlated with the practice of open defecation in India at a statistically significant level (Rahman et al, 2020) and the likelihood of a child suffering from stunting decreases when they access adequate water and sanitation provision (Fink et al, 2011). There is much evidence that exposure to diarrhoeal diseases in early childhood has a severely detrimental effect on cognitive development (Eppig et al, 2010; Pinkerton et al, 2016). This evidence suggests that access to safe WASH services and WASH knowledge could be integral to children's health, safety and development.

The school environment can serve to mitigate WASH-related inequalities and has much potential to play a central role in bringing about progress towards SDG 6 through improvements in children's WASH access. Schools can offer a safe space to children to access clean toilets, uncontaminated water and well-maintained hygiene facilities that they may not access in the home environment, thus reducing children's exposure to life-threatening diseases and improving their health outcomes (Chatterley et al, 2018). Moreover, SDG 4 (Quality Education) is explicitly linked with SDG 6 with Target 4.a aiming to provide "education facilities that are child, disability and gender sensitive and provide safe, non-violent, inclusive and effective learning environments for all" (UNGA, 2015, p. 17). The indicator for this SDG 4 target explicitly mentions access to basic WASH facilities

including potable water, gender-separate toilets and handwashing stations. Thus, the importance of WASH at the centre of education is evident.

Development agencies and non-governmental organisations (NGOs) have exerted much effort to improve school WASH provision. Thorough investment in infrastructure such as toilet blocks and handwashing stations gives children the opportunity to practise positive WASH behaviours such as handwashing with soap and drinking a daily sufficient amount of clean water. The literature shows that such infrastructural upgrades can have a positive impact (Erismann et al, 2017; Chard et al, 2018; Shrestha et al, 2020). However, sustained behaviour change is difficult to achieve and this is the reason why the targets of many schools-based WASH interventions are not successfully achieved in the long term (Humphrey, 2019; McMichael, 2019). Therefore, the current research applies the Capability, Opportunity, Motivation Behaviour (COM-B) model of behaviour change (Michie et al, 2011), a useful framework for evaluating behaviour change interventions, to investigate how schools-based Water, Sanitation and Hygiene (WASH) interventions can sustain children's WASH behaviours. The following sections set out a nationwide example of this issue through the Indian government's Swachh Bharat Abhiyan (Clean India Mission, 2014-2019) campaign before summarising previous studies focusing on sustained behaviour change adoption following schools-based WASH interventions. Literature pertaining to the application of the COM-B model and grounded theory analysis in a Global South context is also presented, demonstrating that this is an appropriate approach for the current research and highlighting how it contributes to the wider research.

Infrastructural renovation without sustained behaviour change

The prioritisation of infrastructural improvements over effort to bring about sustained behaviour change is a particularly urgent problem as a low level of adherence to WASH-related recommended behaviours will undermine the aim of SDG 6 to "Ensure availability and sustainable management of water and sanitation for all" (UNGA, 2015, p. 18). On a national scale, this issue is demonstrated by the Indian government's Swachh Bharat Abhiyan (Clean India Mission, 2014-2019) campaign which was described as "the largest behavioural change programme in the world" (Mohapatra, 2019, p. 451). However, it was criticised for an imbalanced approach which reduced its impact, with an excessive focus on household latrine construction at the expense of participatory behaviour change activities (Jain et al, 2018; Mohapatra, 2019). These two interwoven strands of WASH interventions are referred to in the literature as WASH hardware and software (Darteh et al, 2019; Dey et al, 2019). Because the Swachh Bharat Abhiyan did not include a sufficient WASH software component, its impact was limited and the government's claim that the programme eliminated open

defecation in every state in India was proven untrue by research (Gupta et al, 2019). In relation to Delhi specifically, there is evidence of a similar concentration on WASH hardware without sufficient consideration of WASH software. The Economic Survey of Delhi 2019-2020 (Government of NCT of Delhi, 2020) states that all schools in Delhi have drinking water stations and gender-separate toilets. However, this government report presents an incomplete narrative. These facilities may indeed be present but the extent to which they are maintained, cleaned, functioning and used properly varies greatly. This points to a common challenge in interventions to build WASH capacity: infrastructural changes must be accompanied by successful behaviour change programmes if they are to bring about the targeted long-term outcomes.

Sustaining WASH intervention behaviour change outcomes long-term

The sustainability of behaviour change outcomes is frequently highlighted in the literature as a key issue in evaluations of WASH interventions (Garn et al, 2017; Ejelonu et al, 2020; Orgill-Meyer and Pattanayak, 2020) and thus merits further research. Although not focusing on schools specifically, Martin et al (2018) conducted a systematic review of the literature pertaining to the long-term adoption of WASH-related behaviour and technology, finding that the time frame used in evaluative studies to measure the sustainability of intervention outcomes is often too short. They highlight “the need for more systematic definitions of sustained adoption” (p. 133), recommending that evaluations of WASH interventions should increase the frequency of measurements of behaviour change and incorporate multiple measurement techniques to investigate more accurately to what extent stakeholders adhere to intervention recommendations over time.

Focusing on schools-based WASH, handwashing with soap is a particularly difficult behaviour in which to bring about long-term behaviour change amongst school students (Chard and Freeman; 2018; Humphrey, 2019). However, findings from a large-scale study in Indonesia demonstrate that it is possible; Karon et al (2017) find that intervention schools were more likely to have functioning handwashing with soap facilities than non-intervention school twelve months post-intervention. Bresee et al (2016) even find evidence in Zambia that with appropriate support, children can pass on small handwashing behaviour changes from school-based interventions to their household. However, children should not be burdened with the responsibility for generating WASH behaviour change within their communities (Joshi et al, 2016).

Several studies point to the important role that teachers play in encouraging and modelling the long-term adoption of positive WASH behaviours, especially for younger children (Ghanim et al, 2016; Wagner et al, 2019). Saboori et al (2013) link the varied extent of teachers’ handwashing advocacy in intervention

schools to the observed inconsistency of correct handwashing techniques amongst students. Secondly, Harahap et al (2018) conclude that teachers' influence is paramount in ensuring that students adopt healthy hygiene behaviours in the long term following WASH interventions. The positive correlation between students' hygiene practices and the cleanliness of the school environment is stronger when teachers' behaviour is also taken into account. La Con et al (2017) note a clear improvement in students' hygiene behaviours following a schools-based WASH intervention which included the installation of handwashing and drinking water facilities alongside hygiene training for teachers. Hetherington et al (2017) also identify a key role for teachers. Their WASH intervention in Tanzania included teacher workshops and extra-curricular activities. It resulted in a decrease in participants practising damaging WASH-related behaviours and an increase in participants understanding the vital importance of handwashing with soap. Teachers' engagement with the intervention's participatory methods was cited as particularly key to its success. Finally, de Albuquerque (2014) notes that incorporating community participation into intervention design has the potential to increase the sustainability of outcomes considerably and pave the way for a shift in social attitudes.

In summary, the literature demonstrates evidence that the effectiveness of WASH interventions which do not prioritise the facilitation of sustained behaviour change is limited. There is some evidence that schools-based WASH interventions can succeed in enabling the adoption of positive WASH behaviours such as handwashing with soap. More detailed post-intervention observation of behaviour at regular intervals is required to ascertain the extent of behaviour change. Finally, teachers' engagement with the intervention recommendations is paramount in ensuring that outcomes are sustained in the long term.

Utilising the COM-B model and grounded theory to understand the factors driving WASH behaviour adoption

The present research applies the COM-B model to investigate how schools-based WASH interventions can bring about sustained changes in children's WASH-related behaviours. The Capability, Opportunity, Motivation Behaviour (COM-B) model (Michie et al, 2011) is a theory of behaviour change which can contribute towards meeting these targets by acting as a framework to design and evaluate schools-based WASH interventions. It states that these three domains of capability, opportunity and motivation must be in place for successful behaviour change to occur. Michie et al (2011) describe the model:

"Capability is defined as the individual's psychological and physical capacity to engage in the activity concerned…Motivation is defined as all those brain processes that energize and direct behaviour…It includes habitual processes, emotional responding, as well as analytical decision-making. Opportunity is defined as all the factors that lie

outside the individual that make the behaviour possible or prompt it." (p. 4)

The COM-B model is an appropriate framework to apply to the current research as it has been used extensively in Global South contexts but not in the context of primary schools-based WASH interventions in India. In their systematic review of hygiene interventions in settings occupied by children, Staniford and Schmidtke (2020) include just two India-based interventions, neither of which use the COM-B model (Biran et al, 2014; Lewis et al, 2018). Arriola et al (2020) employ the COM-B model to design a behaviour change intervention to address stunting and increase adoption of nutrition and WASH-related behaviours for pregnant mothers and guardians of children under two in Kenya. The authors find that the COM-B model facilitated the design of a structured, rigorous, theory-based programme which targets specific behaviours but that is not too narrow in scope. McGuinness et al (2020a) conducted a cluster randomised trial investigating the usage of riverbank filtration methods to improve clean water access in rural India (Karnataka). The authors build on this with a qualitative study (McGuinness et al, 2020b) which applies the COM-B model in a similar way to the current research – identifying impeding and enabling factors which affect WASH behaviours. In this case, the behaviours in question were the increased usage of filtered water and the frequency of health reporting. Using the COM-B model and the same qualitative methods as the current research (focus group discussions and semi-structured interviews), the authors identify a number of factors affecting behaviour adoption including knowledge and attitudes towards the links between water and health, convenience, a reluctance to share information and inconsistency in how survey questions are interpreted. The authors conclude by stating that the identification of these enabling and impeding factors will be useful in designing the implementation and evaluation of future interventions. Ellis et al (2020) also aim to identify factors affecting the adoption of nutrition and WASH behaviours. These include faeces disposal, use of latrines and the provision of clean play environments for children under two years old. Employing a mixed-methods design using household observations, focus group discussions and interviews, they apply the COM-B model to identify barriers to behaviour adoption. Barriers to widespread latrine use included the social acceptance of open defecation and the prohibitive cost of durable materials. Factors impeding safe faeces disposal included the scarcity of latrines and embedded beliefs that children's faeces are not unhygienic. Finally, the identified barriers to the provision of clean environments for play were the need for maintenance and the issue that inhabited spaces were shared with animals. Again, the authors conclude by explaining that the use of a defined behaviour change framework is useful in designing interventions founded upon theory and which take these barriers to behaviour adoption into account.

Focusing specifically on the COM-B model in the context of schools-based

WASH interventions, Staniford and Schmidtke's (2020) systematic review points to evidence that applying the multi-strand approach of children's capability, opportunity and motivation to practise positive WASH behaviours is beneficial in achieving the desired outcomes. They argue that since WASH is a complex issue, interventions addressing a single COM-B domain may not be sufficient to produce lasting behaviour change. Each complementary element must be in place to ensure sustained behaviour adoption. For example, Okello et al (2018) use the COM-B model to evaluate a schools-based WASH programme in Tanzanian primary schools which provided WASH software components including teacher-led hygiene education, participatory activities, handwashing demonstrations, engagement with parents and supportive 'nudge' elements such as hands painted on handwashing facilities. The authors investigate barriers and enabling factors for children's handwashing behaviour, finding that children's motivation for handwashing increased as they better understood the negative outcomes associated with not adopting this behaviour. Children had the opportunity to practise handwashing through the provision of facilities but soap and water were not consistently available. The authors conclude by calling for "more theoretically informed research…to unpack the drivers of institutional factors that contribute to behavioural outcomes" (p. 9).

The previous literature shows how the COM-B model can be applied to investigate factors affecting adoption of behaviour during and following WASH interventions and advocates further research in this area. It is also important here to look to the previous literature to explain why the grounded theory analysis approach was employed in the present research. Grounded theory employs the procedure of "constant comparisons" (Corbin and Strauss, 2015, p. 7) in which "Categories are linked and organized by relationship, conditions and dimensions are developed, and finally a theory emerges" (Scott, 2004, p. 113). It is particularly useful when investigating areas where little previous research has been conducted (Engward, 2013), such as the application of the COM-B model in the context of Delhi primary schools-based WASH programmes. Some previous WASH research based in the Global South has used the grounded theory approach to analyse qualitative data. The analysis processes used in three examples are presented here to justify the selection of grounded theory analysis for the current research.

Hennegan et al (2020) investigate women's experiences of menstrual hygiene management in the workplace, commencing grounded theory analysis during the data collection process. They began analysis with systematic coding of each line of data before the authors met to identify common emerging themes and establish an overarching theme: "being a responsible woman" (p. 1). A clear strength of the implementation of grounded theory analysis concurrent with data collection was that the researchers were able to adapt their approach to interviews according to the themes emerging from the data. Mulopo and

Chimbari (2021) also adopted the grounded theory approach, investigating schistosomiasis prevention through WASH strategies in South Africa through qualitative methods including key informant interviews and focus group discussions. The authors applied grounded theory by determining the common themes in the data and categorising them to better understand the factors which increase the risk of schistosomiasis infection. They read the audio transcripts several times before systematically coding the data according to the key themes of their literature review. These themes were then cross-checked with the raw data to ensure the validity of the findings before the themes were given titles for the final report. This methodical, rigorous approach enabled the researchers to identify key categories and subcategories pertaining to the research focus and understand better how key WASH behaviours could assist in reducing schistosomiasis infection.

Finally, Sahoo et al (2015) use the grounded theory approach to analyse in-depth interviews with women in Odisha, India, investigating their experiences of stress related to sanitation. As with the approach of Hennegan et al (2020), analysis was completed alongside data collection, as well as afterwards. The researchers began the grounded theory analysis process by studying initial interview transcripts and preparing "detailed summaries ('memos')" (p. 82). In collaboration, they agreed on emerging themes which informed the adaptation of the interview guides and decisions around sampling. With an iterative approach, systematic coding took place halfway through data collection and themes were further clarified and developed. The researchers used the in-depth summaries to prepare analytical summaries which organised the themes into a coherent structure and finally, these were discussed and compared. The result was a final analysis of the data which revealed valuable insights into women's experiences of sanitation-related stress.

The present research adds to the literature relating to WASH interventions and builds on previous findings which recommend further theory-based investigation in this area. It achieves this by using the grounded theory approach to apply the COM-B model in the context of primary schools-based WASH interventions in India and providing clarity on the extent to which the COM-B model is a useful tool in driving progress towards SDG 6 in a schooling context. The following section outlines the research focus and the themes explored in relation to the COM-B model before the methodology section sets out the research setting, methods and analysis approach employed, as well as ethical considerations.

Research question and themes

In light of the gap in knowledge emerging from the previous literature regarding the application of the COM-B model to primary schools-based WASH programmes, the present research is situated in East New Delhi primary schools and uses grounded

theory analysis of primary qualitative data (semi-structured interviewing and ethnographic observations including photography and a daily reflective diary). It explores each of the three domains of the COM-B model – capability, opportunity and motivation – to understand how intervention outcome adherence could be improved in this context. Specifically, the research question is as follows:

How can schools-based Water, Sanitation and Hygiene (WASH) interventions sustain children's WASH behaviours post-implementation?

The themes explored are:

1. the potential of schools-based WASH interventions to instigate progress towards SDG 6 (opportunity and capability);

2. the reluctance of school staff to play a role in sustaining WASH behaviour change intervention outcomes long-term (motivation); and

3. common barriers to sustained WASH behaviour change adoption (capability).

In light of these themes and the previous literature, the discussion section then explores the potential of the COM-B model to act as a tool for designing schools-based WASH interventions or for identifying the gaps in existing programmes.

Methodology

This methodology section details the schools-based WASH intervention which took place in two of the schools participating in this research, clarifies the different management types of the four schools, and explains why Delhi is an appropriate choice of location and where the schools are located within the region. The remainder of the section describes the qualitative research methods employed, the grounded theory approach adopted during analysis and the ethical considerations.

Research setting

Four East Delhi primary schools participated in this research. Two of the government schools participated in a multi-strand WASH intervention, implemented by an Indian NGO, consisting of both hardware and software elements in 2018. The other two, one government and one low-fee private, did not. The hardware aspect of the intervention included the following components: renovation of toilet blocks, handwashing stations and drinking water facilities; the rejuvenation of schools' green areas; the installation of a rainwater harvesting system to mitigate insecure water connections; and an anaerobic wastewater treatment system to recycle water for nurturing gardens in the school compounds. Meanwhile, the complementary software aspect included participatory activities such as games, art, drama, parent-teacher association meetings, the installation of information posters and a class monitor system which rotated the responsibility around different students.

It is important to clarify the different school management types featured in this research. The Bharatiya Janata Party (BJP) is led by Prime Minister

Narendra Modi and is the national ruling party having won a majority in 2019. The three government primary schools which participated in this research are managed by the East Delhi Municipal Corporation (EDMC) which is governed by the BJP. These are not to be confused with schools managed by the Aam Aadmi Party (AAP), the ruling local government party for the National Capital Territory of Delhi. It manages secondary schools which have received international acclaim following significant investment and improvements through the use of robust accountability measures, including CCTV, to raise outcomes (BBC News, 2018; Biswas, 2020). The fourth school which participated in this research is a low-fee private school. This is a phenomenon common in India and other countries in the Global South which endeavours to offer high-quality, inexpensive education to parents as an alternative to government provision (Mond and Prakash, 2019).

Delhi is a suitable location for schools-based WASH research as it is an area affected by accelerated groundwater depletion and therefore water insecurity, particularly for low-income households (Das, 2020). Although the practice of open defecation has decreased considerably in Delhi in recent years (Rahman et al, 2020), WASH-based inequalities such as varying household water access and inconsistent school WASH provision nonetheless persist.

Two of the six intervention schools (Schools A and B) agreed to take part. It is common for researchers to experience difficulties in obtaining permission to gather data in schools managed by the BJP; teachers are obliged to promise in writing that they will not give interviews to media outlets (Kalra, 2019). Nonetheless, a non-intervention school was successfully recruited (School C) from the list of schools that were due to participate in the intervention in the future. The LFP school (D) was approached following a recommendation from the researcher's contact at the National Independent Schools Alliance (NISA). Both intervention schools (A and B) are situated in Shahdara, a low-income district of East New Delhi. The non-intervention government school (C) is situated in Mustafabad, several kilometres to the north of the intervention schools. All three of these schools run separate shifts for male and female students with boys attending in the morning and girls in the afternoon. This is relevant to the research because the principal of the boys' shift at School B declined to take part so all the School B children who participated in the study are female. Finally, the non-intervention LFP school (D) is also in Shahdara, a few kilometres west of School A, and is co-educational.

Research methods

Qualitative data (semi-structured interviewing and ethnographic observations including photography and a daily reflective diary) were collected in the four participating schools, as well as in the communities surrounding them, with the aim of ascertaining how WASH-related behaviour change following WASH

interventions can be sustained over time. Semi-structured interviewing was employed with the adult participants (4 school principals, 12 teachers and 2 janitorial staff). At Schools A, B and D, this was in the form of recorded audio. At School C, two participating teachers declined to be recorded so this was carried out through the researcher's notetaking during the interview instead. The advantage of semi-structured interviewing is that while the researcher formulates a set of questions, the interview is conducted in an informal, casual manner so that participants have the opportunity to broach subjects that they want to discuss (Longhurst, 2016) and the researcher is able to situate their understanding within the participant's perspective (Patton, 1987). Ethnographic observations were also conducted in the schools, including photography of school infrastructure and a daily reflective diary informed by fieldwork notes from environmental observations and interactions with stakeholders (Bryman, 2012). For ethical purposes, the photography only included the school premises and no participants appeared in photographs.

Although ethnography usually occurs in a long time frame, this was not viable due to the limited time available for data collection. Thus, the approach followed was similar to rapid ethnographic assessment which is employed "to quickly create a socio-cultural profile of a group or region" (Carley et al, 2012, p. 300). While not comprehensive to the same degree as rapid ethnographic assessment, these observations are nonetheless valuable in complementing and validating the data from the interviews and focus group discussions which present the voices of stakeholders.

There are some limitations involved in employing these methods and it is important to report these which constitute "the systematic bias that the researcher did not or could not control and which could inappropriately affect the results" (Price and Murnan, 2004, p. 66). First, with regards to semi-structured interviews, some people may not be comfortable talking about sensitive topics such as open defecation. This is perhaps especially true of school staff who understandably do not wish to portray their employers in a negative light. Thus, social desirability response bias, which is "a tendency to present reality to align with what is perceived to be socially acceptable" (Bergen and Labonté, p. 783), may have affected the data. This was mitigated as much as possible by including a variety of different stakeholders' perspectives and thus triangulating data sources. Regarding ethnographic observations and photography, it is important to recognise that these were collected through the lens of the researcher's decisions relating to details and spaces in the environment important to the research aim. As such, the researcher's perspective in the data colours the representation of these school environments (Malterud, 2001). Furthermore, due to ethical considerations, it was impossible to photograph or record children's WASH behaviours (e.g. handwashing techniques, usage of drinking water facilities). This would have offered rich insight into how the WASH intervention has affected their behaviours long-term.

Analysis

Each of the principal stages of the grounded theory approach was adopted during analysis: initial coding, analysis, memo writing and theoretical categorisation (or selective coding) (Glaser, 1992; Charmaz, 2015). The qualitative data analysis software package NVivo supported this approach; textual data were imported into NVivo and audio data were transcribed before being imported. The use of NVivo aligns with the grounded theory approach because it assists the researcher in "identifying significant patterns…drawing meaning from data and subsequently building a logical chain of evidence" (Wong, 2008, p. 14) and enables "the exploration of values, meanings, beliefs, thoughts, experiences, and feelings" (ibid.).

Grounded theory uses inductive reasoning, "the process of moving from the specific to the general" (Sandberg and McCullough, p. 179), to build a theory according to the data, rather than commencing the process with a presupposed theory already in place. As such, it is a useful approach when little is known initially about the research focus (Engward, 2013). As with the approach of Hennegan et al (2020) and Sahoo et al (2015), the application of grounded theory analysis to the current research began during data collection with initial coding. Following each school visit, the rough ethnographic observation notes were written up and synthesised into a reflective diary, organised into common themes emerging from data collection on each given day, rather than in the order of when the interviews took place. The researcher's reflections on these emerging themes were added below the synthesised data to connect the process of coding to the emergence of theory (Charmaz, 2015). This meant that an iterative process of reflection, similar to the group approach of "detailed summaries ('memos')" adopted by Sahoo et al (2015, p. 82), was followed on a daily basis during data collection and informed the direction taken in subsequent interviews with school staff.

Following data collection and transcription of audio recordings, the researcher read, re-read and systematically coded each line of textual data (interview transcriptions and the daily reflective diary), applying the concept of "constant comparisons" (Corbin and Strauss, 2015, p. 7). In the process of coding, each part pertaining to school WASH facilities, WASH behaviours, behaviour change and the impacts of the WASH intervention was highlighted, according to the research focus. The photographs were reviewed and those relevant were retained and coded conforming to the same themes. The researcher then reviewed these coded textual data and photographs, aiming to identify common categories shared across the different data sources. Concurrently alongside this initial coding process, the researcher used memo writing to consider how these common categories relate to one another to reveal a deeper understanding of how schools-based WASH interventions can sustain children's WASH behaviour adoption, writing a plan of how the final report would fit together. Following the same approach as Mulopo and Chimbari

(2021), this also enabled the researcher to cross-reference initial findings, interrogating assumptions to ensure that they were grounded in the data. Thus, data triangulation, defined as the "use of more than one method or source of data in the study of a social phenomenon so that findings may be cross-checked" (Bryman, 2012, p. 717), was employed to identify robust, clarified findings in the data. In the context of this study, different stakeholders in the schools (principals, teachers and janitorial staff) offer a range of perspectives which, when triangulated, construct a detailed understanding of the phenomena affecting the sustainability of the schools-based WASH intervention's behaviour change outcomes.

Through this process, the initial categories were then coded more selectively according to their common themes that arose during memo writing and how these shed light on the research focus. These final core themes were the following and constitute the main findings of the research: the positive impacts of the WASH intervention in Schools A-B in comparison to the non-intervention government School C, the dependence of school staff on the NGO to bring about WASH-related change and barriers to the long-term adoption of WASH behaviour change. These categories emerging from the application of grounded theory form the findings of this article and are situated in the wider landscape of the previous literature in the discussion section.

Aligning with the recommendations of Cohen et al (2018), member checking and peer debriefing were employed alongside triangulation to increase the validity of the data. Member checking "consists of taking data and interpretations back to the participants…so that they can confirm the credibility of the information" (Creswell and Miller, 2000, p. 127). The researcher was not in contact with participants following data collection so the approach of member checking was adapted to take place during this. The main points of comments were repeated to participants for them to clarify, especially where the meaning was ambiguous. Peer debriefing is a process by which an impartial colleague offers critical feedback on research design and data analysis (Lincoln and Guba, 1985). This was achieved through the input of two colleagues at Newcastle University working in the same field.

Ethics

Ethical approval for this research was granted by the Newcastle University School of Education, Communication and Language Sciences Ethics Committee. The British Educational Research Association's guidelines pertaining to recruiting participants, conducting research in a cross-cultural environment, risk assessment, voluntary informed consent, incentives for participants and data storage were followed (BERA, 2018). The researcher gained access to the participating government schools through the local NGO which implemented the intervention. A memorandum of understanding was

established between the NGO and the EDMC, as well as between the researcher and the NGO. Households were recruited for parent interviews through a member of staff from the NGO and the principals of the participating schools.

Two interpreters, both native to Delhi, assisted the researcher during data collection which took place in Hindi. One of these also led the translation of the research materials in advance of the data collection period. A full risk assessment was conducted and all necessary precautions were taken to protect all involved. Voluntary informed consent was required from all participants before they took part in any aspect of the research. Participant information forms were provided which included details relating to the aims of the research, what was involved, the length of time required and participants' right to withdraw their consent. These were read aloud by the interpreters due to the varying literacy levels of children and parents. No incentives were offered to adult participants for taking part but on completion of their involvement, children were given a 'Water is Life' sticker and a certificate as a small token of gratitude. During the analysis process, all necessary precautions were taken regarding the storage of digital and hard data.

Results

Here, the core themes emerging from the application of the grounded theory approach to the data are presented in response to the research question. Each theme corresponds to one of the three domains of the Capability, Opportunity, Motivation Behaviour (COM-B) model. The first theme revealed during the analysis process is the clear potential for schools-based WASH interventions to provide the opportunity for children to adopt safe and hygienic WASH-related behaviours and thus bring about progress towards SDG 6. The improvement in infrastructure brought about by the intervention is illustrated through an outline of WASH facilities issues observed at the non-intervention School C that were not observed at the intervention Schools A and B. In addition, insight from school staff regarding the impacts of the intervention reveal how it also built capability for children to develop positive WASH behaviours through increased school attendance. The second core theme emerging from the grounded theory analysis process relates to the COM-B model domain of motivation and considers the reluctance of school staff to play a role in sustaining WASH behaviour change outcomes post-intervention. The third builds on this, relating to common barriers to sustained WASH behaviour change adoption. These are teachers' obligatory additional commitments, the motivation of staff to fulfil school duties and the high turnover of students enrolled in the schools. Each of these three common barriers speaks into the domain of capability.

Schools-based WASH interventions: a potentially powerful tool in progress towards SDG 6

The opportunity domain of the COM-B model refers to "all the factors that lie

outside the individual that make the behaviour possible or prompt it" (Michie et al, 2011, p. 4). During the coding process of "constant comparisons" (Corbin and Strauss, 2015, p. 7), contrasts between the three participating government schools emerged from the data. These indicated that some infrastructural factors at School C (non-intervention) were in fact denying children the opportunity to practise positive WASH behaviours, particularly in relation to the toilet blocks. These issues were not present to the same extent at Schools A and B (intervention). They were outlined by two teachers from School C in an interview who did not consent to be recorded. First, they cited the insufficient janitorial staffing at the school with just one member of cleaning staff for a school with approximately 1,300 students. This issue combined with inadequate drainage systems in the toilet facilities resulted in an unsafe and unclean environment (Figure 1). The tap in the photograph was not functioning during each visit. Secondly, the school's supply of potable water was compromised because the reverse-osmosis purification system was not functioning.

Figure 1. Male students' toilets (School C 10th February 2020, 13.20)

As these facilities did not give students the opportunity to practise healthy behaviours throughout the school day such as handwashing with soap and drinking clean water, they had to rely on bringing water from home. This negated the role the school environment could potentially play in addressing household WASH inequalities. Furthermore, the insufficient number of handwashing stations in the school meant that children were habituated to washing their hands by pouring their water bottles on their hands outside the classroom windows. Not only was this ineffective without soap but also wasted the children's drinking water that they had brought from home. In comparison, the drinking water facilities at School B were functioning well and children, many of whom experience daily water insecurity at

home, were observed filling plastic reusable bottles during breaktime.

Another factor impeding children's opportunity to practise hygiene behaviours observed at School C (non-intervention) could particularly affect girls, thus demonstrating how progress towards SDG 6 (Clean Water and Sanitation) is intrinsically linked with SDG 5 (Gender Equality). A ladder was propped up against the upper window of the girls' toilets during each visit conducted by the researcher (Figure 2). Identical windows above toilet cubicles inside meant that anyone could illicitly observe students using the toilets. No female students were observed entering these facilities during the researcher's visits to School C. In comparison, the toilets at Schools A and B were located within the main school building with windows for ventilation which did not compromise students' privacy. In fact, according to a member of staff from the NGO which implemented the intervention, this was prioritised in the toilet block renovation design after a female student raised the issue in preliminary stakeholder focus group discussions.

Figure 2. Student toilet blocks (School C 10th February 2020, 13.20)

In addition to these observations, the application of grounded theory highlighted common positive opinions shared by staff from Schools A and B regarding the work of the NGO in their respective schools. These contribute further evidence to a potential key role for schools-based WASH interventions in furthering progress towards SDG 6. In the COM-B model, capability refers to "the individual's psychological and physical capacity to engage in the activity concerned" (Michie et al, 2011, p. 4). In other words, it concerns the knowledge and skills that are required to engage in the behaviour. One key aspect of ensuring that children have the capability to adopt healthy WASH behaviours is their school attendance. The principal of School A indicated that the WASH intervention had brought about an increase in this:

"[NGO] people went to visit the parents of students and through various activities they have connected with them. They have worked in the communities and collected students to

bring them to school…They have started various activities – students are more excited and encouraged specifically for these activities. So [NGO's] work in this way increased attendance because students were more keen to be here. Students are happy to work with activities like cleaning or planting a tree."

"Students go with the parents to earn money begging, to work on the farms. They miss school and this makes attendance lower…One of the main things is [NGO] acts as an interface between the school and the community. So the dropouts from school, [NGO] brings them back." (School A girls' principal, 16th January 2020)

In contrast with the barriers girls face in sanitation privacy at School C, one teacher from School B also linked the WASH intervention to an improvement in support for female pupils' menstrual hygiene management and subsequent increase in attendance:

"Girl children were taught especially how to take care during menstruation. The students were not aware, what happens, how they can keep clean during this themselves. During those days, girl students didn't come in, they used to remain at home…Only a few girl students have access to menstrual hygiene products, not all of them…They do not fear about it now, they take it normally now as a natural process and they come to school." (Female teacher at School B, 16th January 2020)

Here, capability and opportunity complement one another as girls' access to safe, private spaces (opportunity) appeared to increase their school attendance and thus their capability or "psychological and physical capacity to engage in the activity concerned" (Michie et al, 2011, p. 4).

School staff's dependence on the NGO to sustain WASH behaviour change intervention outcomes long-term

Another School B teacher expressed her hope that the children's learning from the NGO's behaviour change communication activities would result in sustained adoption of positive WASH-related behaviours:

"We hope that the children will carry these changes with them for a long period of time. These changes can just be for a limited time but they should remain for a long time." (Female teacher at School B, 16th January 2020)

However, in the main, staff did not show willingness to uphold and sustain behaviour change outcomes post-intervention, suggesting that the motivation element, "defined as all those brain processes that energize and direct behaviour" (Michie et al, 2011, p. 4), was not in place on the part of the teachers. Comments from staff indicate a dependence on the NGO's efforts to effect immediate change without a commitment to ensure the continuation of this in the long term. This is despite the fact that staff at Schools A and B expressed positive sentiments regarding the intervention. For example, one teacher described the NGO's work in the community surrounding the school:

"Illness has decreased and the reason is because [NGO] works with the community as well. [NGO] educates the parents in the community. The cleanliness awareness maybe decreased the case of illness. Attendance is rising." (Female teacher at School A, 16th January 2020)

However, the NGO's work had ended and there was no indication in her comments that school staff could continue to build these important community links. In addition, another teacher at School A pointed to the fact that children need repeated reminders of positive WASH behaviours but did not offer suggestions as to how school staff could fulfil this role following the end of the NGO's formal agreement with the school:

"The infrastructure, the activities performed by the [NGO] people. These are things regularly the children want. Every day, they need such activities, they are children. You tell them every day, do this thing, do this thing then they don't do it the next day. So regularly you should tell them. Yes these are the very small children – they need to be taken care of regularly. They are not much capable – that thing is taught once. Every day we need to teach them." (Male teacher at School A, 17th January 2020)

This raises questions as to why the WASH intervention behaviour change outcomes were not sustained to the extent that was intended during the design phase. The following section explores this, outlining three barriers to children's sustained adoption of WASH behaviour change post-intervention. These are teachers' obligatory additional commitments, the motivation of staff to fulfil their school duties and the transient nature of students' enrolment in the schools. Each of these relates to the COM-B model domain of capability.

Barriers to children's sustained adoption of positive WASH behaviours

The previous section highlighted some infrastructural barriers denying children the opportunity to adopt healthy hygiene behaviours at the non-intervention government school (C) which were not present to the same extent at the intervention government schools (A and B). This indicates that the intervention laid the foundation for enduring positive outcomes through sustained behaviour change and even increased children's capability to adopt the targeted behaviour through increased school attendance. School staff members' comments about the intervention reinforced this, as well as indicating that they do not have the motivation to support children in sustaining positive behaviour change long-term.

During the grounded theory analysis process, three reasons emerged from the different data sources as to why the targeted behaviour changes were not maintained over time, despite the various successes of the schools-based WASH intervention. These are the additional commitments imposed upon teachers, school staff's motivation to discharge their school duties and the high turnover of students attending the schools. Each of these applies to children's capability to practise positive WASH behaviours, their "psychological and physical

capacity to engage in the activity concerned" (Michie et al, 2011, p. 4). These three reasons are particularly relevant to the research question because if children's capability is enhanced through the increase in school attendance brought about by the intervention, this is negated if they do not access adequate support in the school environment.

During the researcher's visits to Schools A-C, teachers were often absent as a result of their additional commitments as 'Block Level Officers' in local elections (Mitul and Agha, 2018). The schools were not provided with substitute teachers when the regular staff were fulfilling these extraneous duties, meaning that full classes of children were left unsupervised for the entire school shift. Chaitanya MRSK, an education activist in India, reports in Education World (2018) on the problem of overloading teachers with additional responsibilities:

"The drawback is that the government looks at these teachers as government employees. This attitude needs to change. We need to look at teachers in government schools as agents of social change. Yes, it is true that there is a lot of burden on the teachers to take up administrative work besides teaching. This does hamper their quality of teaching." (para. 5)

If teachers are required to fulfil duties unrelated to teaching which require them to miss time in the classroom, they are unable to support children in developing the capability to practise positive WASH behaviours because they have fewer opportunities to act as role models and provide health education. Moreover, it is also reasonable to suggest that they would have less motivation to uphold WASH intervention recommendations long-term if their focus is distracted by non-school commitments. The comments from School A staff above intimate that the intervention had brought about an increase in school attendance. It may be true that student absenteeism decreased following the WASH intervention, as these staff members suggest. However, if students who are now coming to school more often are left without a teacher for long periods, it will be difficult for this increase in attendance to translate into a positive impact on children's hygiene behaviours, and on their learning, and their health and safety could be jeopardised. Furthermore, there is a danger that if teachers continue to be burdened with these obligations unrelated to their job, missing lesson time will be normalised which could decrease their motivation to execute their professional duties. This constitutes the second reason for why the behaviour changes targeted by the WASH intervention were not sustained over time.

Another reason emerging from the "constant comparisons" (Corbin and Strauss, 2015, p. 7) analysis process for why behaviour change targets were not sustained long-term is that some staff appeared to lack motivation, or the "brain processes that energize and direct behaviour" (Michie et al, 2011, p. 4), to meet the demands of their role. This has implications for children's capability to practise positive WASH behaviours as they require support and role models to do so. One instance of this was observed by the

researcher when a significant amount of noise was coming from a classroom full of students. The researcher and interpreter entered the classroom to investigate and found that no teacher was present; over ten children were also playing outside the classroom in the corridor. The interpreter asked one of them where the teacher was and what they were doing and the child responded:

"He has not come today…we are going to play with a ball. When he comes, he doesn't teach, he just writes in his book. He doesn't write on the board." (School A male student, 20th January 2020)

On another occasion, the researcher observed two male teachers leaving School A immediately after the end of the assembly which commenced the school session. They had not returned over an hour later when the researcher left. In fact, the interpreter, who coincidentally lived in the same area as a child, suggested that it is possible that they supplemented their government salaries with income from private businesses, perhaps working as tutors. It is important to state that it is conceivable that they were attending their compulsory roles as election officers, as discussed in the above section. However, the interpreter's reasoning was substantiated by a comment from a parent whose four children attend School B:

"They are not getting proper education in the school. The teachers do not teach the students properly. They give their tasks to young girls." (School B mother, 14th January)

Clearly delineated disciplinary procedures are required to ensure that teachers who practise such behaviour are held to account. However, it is important to stress that this accountability must be reciprocated in terms of school management bodies' responsibilities to their staff. This is clearly seen in the current situation regarding primary schools in New Delhi. In March 2020, all primary schools in India closed due to COVID-19 and as of June 2021, those in New Delhi remain closed with the virus continuing to spread (National Herald, 2021). In February 2021, EDMC teachers began indefinite strike action after having not received their salaries for three months (Hindustan Times, 2021). It is also important to note that this observed lack of motivation among teachers was not universal across the intervention schools. For example, a teacher at School B explained her enthusiasm for innovative pedagogical approaches in her efforts to implement the NGO's class monitor system:

"Peer group learning is the best way to teach. And this is what we see – how peers learn between themselves. So when the individual is a monitor and the others follow them, then they learn easily. Because children are teaching them and they are set a good example. Peer group learning is the most amazing learning process." (Female teacher at School B, 16th January 2020)

The NGO did include elements in the intervention to support teachers; a professional development event was organised for the teachers from the intervention schools. A teacher from School A described the training where

they received resources and materials for use in their schools including games and books. However, further regular training on behaviour change for teachers would be beneficial in building their motivation to support children's capability for sustained WASH behaviour adoption long-term. For example, at School D (low-fee private), each class was supervised queuing outside the toilets before lunchtime to ensure handwashing before eating but this would only be possible at Schools A-C with the necessary number of staff members present.

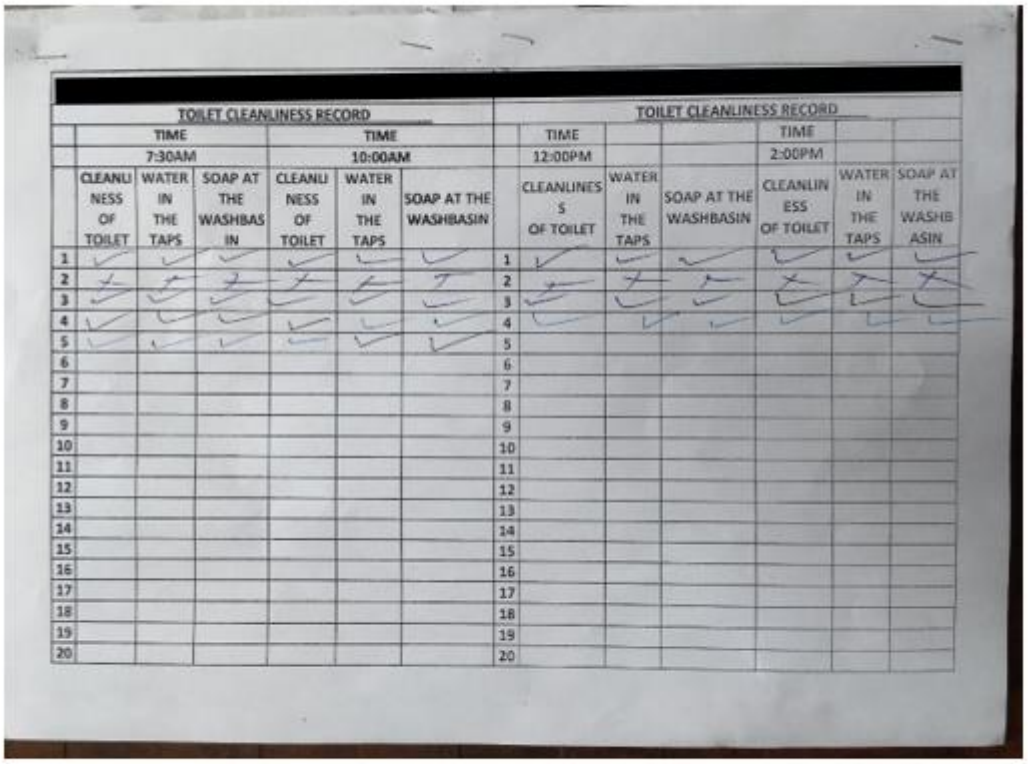
Staff accountability could be achieved through draconian approaches such as classroom CCTV which was effective in the schools managed by the Aam Aadmi Party (BBC News, 2018). However, it would be more beneficial in the long term to build motivation through the provision of frequent training sessions for teachers on the significant role they can play in children's development of positive WASH-related behaviours, using more cooperative methods aiming to enhance their professional development. These could include reflective journals and observations, both from teacher peers and senior management (scheduled and unscheduled), followed by constructive feedback.

In addition to this though, robust accountability measures are necessary to support the maintenance of WASH facilities. There was some evidence of such procedures at School D. The cleanliness of the toilet blocks is evaluated regularly (Figure 3) by the head of the cleaning staff and one of the teachers explained the process:

"The maid in charge does rounds of the school. She is responsible. Sometimes it happens that some area is not clean, she would enquire with the other cleaning staff why is it not clean." (Female teacher at School D, 4th February 2020)

This close monitoring ensured that the students had consistent access to "education facilities that…provide safe, non-violent, inclusive and effective learning environments" as per SDG 4 ((UNGA, 2015, p. 17). In comparison, the janitor from School A explained that he was the one member of cleaning staff for the morning girls' shift at School A and was the only person responsible for ensuring four toilet blocks and thirty classrooms were maintained to an adequate standard during that time. His colleague who was in charge of the afternoon shift was not directly employed by the school but rather was recruited on a short-term basis through an agency. Furthermore, this temporary member of staff was frequently replaced by someone else thus there was no accountability system in place to ensure facilities of the same standard of cleanliness as those at School D. The implementation of the system from School D in Schools A-C, alongside the recruitment of an adequate number of well-resourced, permanent cleaning staff would ensure that hardware installed during the WASH intervention is safe for children to use in the long term. The insufficient number of cleaning staff was particularly observable at School C where there

was just one janitor and 800 students enrolled in the boys' shift. According to the deputy principal, the guidelines state a minimum ratio of 1:300. The person in question was not in attendance each day the researcher visited School C and was covered by two agency workers in an apparently similar arrangement to that in place at School A, again highlighting the need for a sufficient number of permanent staff.



TOILET CLEANLINESS RECORD

	TIME 7:30AM			TIME 10:00AM		
	CLEANLINESS OF TOILET	WATER IN THE TAPS	SOAP AT THE WASHBASIN	CLEANLINESS OF TOILET	WATER IN THE TAPS	SOAP AT THE WASHBASIN
1	✓	✓	✓	✓	—	✓
2	✓	✓	✓	✓	✓	✓
3	✓	✓	✓	✓	✓	✓
4	✓	✓	✓	✓	✓	✓
5	✓	✓	✓	✓	✓	✓
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

TOILET CLEANLINESS RECORD

	TIME 12:00PM			TIME 2:00PM		
	CLEANLINESS OF TOILET	WATER IN THE TAPS	SOAP AT THE WASHBASIN	CLEANLINESS OF TOILET	WATER IN THE TAPS	SOAP AT THE WASHBASIN
1	✓	—	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✓
3	✓	✓	✓	✓	✓	✓
4	✓	✓	✓	✓	✓	✓
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

Figure 3. Toilet cleanliness record at (School D 5th February 2020, 12.25)

The third reason why behaviour change outcomes were not sustained in the long-term is the high turnover of students. If children do not attend school regularly, they are unable to develop the capability, opportunity and motivation to engage in the positive WASH-related behaviours that were targeted by the intervention. For example, according to an NGO staff member, approximately 1,200 children are registered at School B but in reality, only 700-800 attend regularly. This was attributed to the phenomenon of migrant families visiting Delhi especially for festivals before travelling back to their home regions. Secondly, the data revealed that child labour is a common cause of student absenteeism in the participating schools. One member of staff at School C explained that she was a specialist teacher for children joining education for the first time:

"They are brought here to work as child labourers and we enrol them to stop them working as labourers. We have to convince the parents. At first,

the parents will say no because they are there for a specific kind of work. If they do not agree with us, we will tell them about the Right to Education law…We convince the parents that the child will be in safe hands with us and will receive everything like a midday meal and there will be no fees for the education." (Female teacher at School C, 4th February 2020)

This was corroborated by one child who explained her hazardous work as a 'rag picker' (Agarwalla et al, 2017) retrieving recyclable materials to sell:

"I go to the school. We work as rag pickers, I think it is a problem for children that we work. My mother thinks it's fine and after the work, we attend the school." (School B female student, 14th January 2020)

These data relating to child labour provide some insight and context for the high turnover of students at these schools. A third reason is the simple fact that some younger children joined school for the first time after the end of the intervention. As a result of these reasons, many students attending the intervention schools did not benefit from the intervention software inputs (participatory behaviour change communication activities) which occurred before they joined. Three members of staff from School A highlighted the challenges that arose before children accessed the behaviour change activities:

"They would use the toilet but they didn't know how to use the toilet. They would just shit everywhere." (School A girls' principal, 16th January 2020)

"[NGO] people told the students how to sit on the toilet seats and use it properly. Before that, children would go to the toilet anywhere in the toilet block." (School A male janitor, 16th January 2020)

"[NGO] has done a lot of work in this school and they give good qualities to the students like handwashing before eating and after toilet. Students who come to this school didn't know how to sit on a toilet seat because in their locality there is no facility. So they usually defecate in the open. [NGO] specifically taught them how to sit on the toilet seat. Students in the school defecate in the toilet block but not on the seat. It causes problems like bad smell and filthiness in school toilets. The students were taught how to sit on the toilet seat and how to wash their hands properly after the toilet." (Female teacher at School A, 16th January 2020)

It is likely that children who missed these activities will continue practising damaging hygiene behaviours such as defecating on the floor of the toilet block rather than in the toilet itself. There was some evidence of this at School B where the researcher found faeces on the corridor floor outside the toilets; the janitor explained that this was likely left by the younger children who had joined the school post-intervention. This indicates that the targeted behaviour change outcomes were not sustained beyond the students who experienced the behaviour change communication activities first-hand. It also returns to the research question: how can schools-based Water, Sanitation and Hygiene (WASH) interventions sustain children's WASH behaviours post-implementation? The following

discussion section considers this in light of the results from analysis and in relation to the existing literature.

Discussion

Three core themes, each corresponding to the domains of the Capability, Opportunity, Motivation Behaviour (COM-B) model, emerged from the grounded theory analysis of the data. The first is the potential for schools-based WASH interventions to provide children with the opportunity and capability to practise positive WASH-related behaviours. The second is the lack of motivation amongst school staff to assist in sustaining WASH behaviour change outcomes long-term. The third relates to the common barriers to sustained WASH behaviour change adoption faced in the participating schools. This section brings these together in light of the COM-B model and the wider literature on schools-based WASH interventions, using previous research to explore how the benefits from schools-based WASH interventions could be maximised.

Securing long-term adoption of positive WASH behaviours

The COM-B model is a useful framework for identifying factors which enable or impede the sustained adoption of positive WASH behaviours (Arriola et al, 2020; Ellis et al, 2020; McGuinness et al, 2020b). Applying it to this current research, children at the intervention schools received the capability and opportunity to practise healthy hygiene behaviours through the infrastructure and hygiene education elements of the WASH intervention that was implemented in their school, as was the case with Okello et al (2018). However, if their motivation is lacking, the desired behaviours cannot be sustained (Staniford and Schmidtke, 2020). Bresee et al (2016) argue that children need a supportive framework of structured activities to maintain positive behaviour change and even transfer it to their parents.

The literature reveals that teachers can play a key role in this (Saboori et al, 2013; Hetherington et al, 2017; La Con et al, 2017). Equipping teachers to fulfil each strand of the COM-B model, but particularly the motivation strand, could be the defining factor in ensuring lasting beneficial outcomes for schools-based WASH interventions, specifically addressing the issue of children without WASH-based understanding joining the school post-intervention. Teachers could perform this role by running permanent extra-curricular clubs focusing on encouraging WASH behaviours through participatory activities similar to the behaviour change communication that took place at the intervention schools during the NGO's involvement. Alternatively, education officials could build this content into the formal curriculum to make WASH learning compulsory.

As Harahap et al (2018) show, teachers can also play a key role in children's adoption of hygiene behaviours by modelling best practice in school, leading efforts to display WASH education materials such as handwashing posters in the school

building or practising activities with their class such as mandatory group handwashing before meals. If teachers have the motivation to prioritise and champion WASH behaviours in their interactions with children, this could produce a positive effect on children's adoption of healthy hygiene practices, even if they were not enrolled in the school when the intervention took place. In fact, Ghanim et al (2016) conclude that teachers are children's second most significant source of hygiene understanding after parents and advocate an increase in hygiene education for primary-aged children. Wagner et al (2019) even state the case for WASH education for pre-school-aged children, accompanied by training for teachers.

Enabling and equipping teachers to support students in adopting positive WASH behaviours can contribute towards securing long-term outcomes of schools-based WASH interventions. However, this is only possible if teachers themselves are motivated to do so. Thus, the COM-B model can be applied to the behaviour of school staff as well as students. Teachers may have the capability (eg. training and knowledge) and opportunity (eg. employment and salary) to support students' adoption of WASH behaviours but this research shows that these two domains are not necessarily accompanied by the motivation required. Humphrey (2019) explains that "One reason for the poor efficacy of low-cost WASH interventions is their requirement for high user adherence to consistent sustained behaviour change" (p. 1158) and this can be applied to teachers in schools where WASH interventions take place. Garn et al (2017) describe weak adherence as "low level of engagement by beneficiaries" (p. 984) and advocate following up WASH interventions to reinforce outcomes post-implementation. As it is not right for children to shoulder the burden for WASH behaviour change in their communities (Joshi et al, 2016), intervention providers should focus on engaging school staff regularly during and after implementation to maximise positive WASH behaviour adoption. This should be supported and evaluated by the longitudinal, multi-observation methods recommended by Martin et al (2018).

Limitations

There were a number of limitations confronted during this research. First, the researcher's opportunity to undertake observations and interviews at Schools B and C was limited. This is particularly true of School B with the principal of the boys' shift declining to take part. As a result, some data (eg. photographs of facilities) were collected at some schools and not others. Finally, as the vast majority of focus group discussions and interviews were conducted in Hindi with two interpreters native to Delhi assisting the researcher, only the elements of participants' responses which were translated into English by the interpreters were included during the transcription process. As participants often spoke in detail before the interpreter had the opportunity to translate what they had said, this introduced a limitation into the analysis as it is probable that some of the data were excluded as a result.

This process is not atypical however; according to Halai (2007), “”transmuted texts,”…[which] reflect the original, but have been recreated” (p. 344) are a valid form for such qualitative data to take. The researcher relied on these English translations from the recordings due to a lack of resources to fund a translator during the data transcription process.

Conclusion

Progress towards SDG 6 is urgent, with health inequalities worldwide caused and exacerbated by WASH insecurity in low-income communities. This is particularly true in India where nationwide efforts to build WASH infrastructure are not sufficient to bring about the shift in behaviour and social attitudes which will eliminate these inequalities. There is potential that the school environment can play a key role in narrowing the WASH gap by providing a safe and secure space for children to access WASH facilities and acquire WASH knowledge that they do not have in the household environment. This is reflected in SDG 4 (Quality Education) which highlights school WASH facilities as a key component of “non-violent, inclusive and effective learning environments” (UNGA, 2015, p. 17).

As a result, many actors conduct schools-based WASH interventions which, as with the Swachh Bharat Abhiyan national WASH campaign, often provide physical WASH infrastructure such as renovated toilet blocks without bringing about enduring WASH-related behaviour change. The literature reveals that teachers can play a central role in addressing this. Informed by primary research conducted in four East Delhi primary schools in January-February 2020, this article has applied the Capability, Opportunity, Motivation Behaviour (COM-B) model to consider how children’s behaviours related to WASH can be sustained following schools-based WASH interventions. Through grounded theory analysis of primary qualitative data, it has highlighted three themes which inform this research aim. First, schools-based WASH interventions have strong potential to drive progress towards SDG 4 and SDG 6 through enabling children’s access to each of these three domains which constitute behaviour change. In particular, in the intervention schools which served as the focus of this research, there was clear evidence of children’s capability and opportunity to adopt positive WASH behaviours. Secondly, school staff have a key role to play in supporting children in sustained practice of these behaviours. However, this research points to a lack of motivation amongst teachers to uphold the intervention recommendations long-term. Finally, this research highlights three barriers to children’s sustained adoption of WASH behaviours post-intervention: teachers’ obligations in election processes, their apparent unwillingness to fulfil their professional duties and the transient, fluctuating nature of enrolled students’ attendance.

The COM-B model of behaviour change (Michie et al, 2011) offers a clear framework for designing effective schools-based WASH interventions. Teachers may have the capability to

encourage students' adoption of positive WASH-related behaviours. However, to have the opportunity to do so, they must be relieved of the extraneous duties which impede their ability to fulfil their obligations to their students. For them to have the motivation to do so, robust accountability procedures and transparent disciplinary measures are required to ensure that teachers who wilfully neglect their duties face appropriate consequences. Frequent, structured teacher training on WASH-related behaviour change communication is also required to assist in developing teachers' intrinsic motivation to fulfil their role. In this way, they could visualise the profound positive impact they can bring about in children's lives through adhering to WASH intervention recommendations long-term and modelling positive hygienic behaviours.

References

Agarwalla, R., Pathak, R., Singh, M., Islam, F. and Parashar, M. (2017) 'Effectiveness of awareness package on occupational health hazards among ragpickers of New Delhi, India', *Indian Journal of Occupational and Environmental Medicine*, 21(2), pp. 89-93.

Arriola, K.R.J., Ellis, A., Webb-Girard, A., Ogutu, E.A., McClintic, E., Caruso, B. and Freeman, M.C. (2020). 'Designing integrated interventions to improve nutrition and WASH behaviors in Kenya', *Pilot and Feasibility Studies*, 6(1).

BBC News (2018) *Delhi parents to watch government schools via CCTV*. Available at: https://www.bbc.co.uk/news/world-asia-india-42718058 (Accessed: 17 June 2020).

Bergen, N.L., R. (2020) '"Everything is perfect, and we have no problems": detecting and limiting social desirability bias in qualitative research', *Qualitative Health Research*, 30(5), pp. 783-792.

Biran, A., Schmidt, W.P., Varadharajan, K.S., Rajaraman, D., Kumar, R., Greenland, K., Gopalan, B., Aunger, R. and Curtis, V. (2014) 'Effect of a behaviour-change intervention on handwashing with soap in India (SuperAmma): a cluster-randomised trial', *The Lancet Global Health*, 2(3), pp. 145-154.

Biswas, S. (2020) *Why schools are at the centre of Delhi polls*. Available at: https://www.bbc.co.uk/news/world-asia-india-51353525 (Accessed: 15 June 2020).

Brauer, M., Zhao, J.T., Bennitt, F.B. and Stanaway, J.D. (2020) 'Global access to handwashing: Implications for COVID-19 control in low-income countries', *Environmental Health Perspectives*, 128(5).

Bresee, S., Caruso, B.A., Sales, J., Lupele, J. and Freeman, M.C. (2016) ''A child is also a teacher': exploring the potential for children as change agents in the context of a school-based WASH intervention in rural Eastern Zambia', *Health Education Research*, 31(4), pp. 521-534.

British Educational Research Association (BERA) (2018) *Ethical Guidelines for Educational Research (4th Edition)*. London: (BERA), B.E.R.A.

Bryman, A. (2012) *Social Research Methods (4th Edition)*. 4th edn. Oxford: Oxford University Press.

Carley, K.M., Bigrigg, M.W. and Diallo, B. (2012) 'Data-to-model: A mixed initiative approach for rapid ethnographic assessment', *Computational and Mathematical Organization Theory*, 18(3), pp. 300-327.

Chard, A.N. and Freeman, M.C. (2018) 'Design, intervention fidelity, and behavioral outcomes of a school-based water, sanitation, and hygiene cluster-randomized trial in Laos', *International Journal of Environmental Research and Public Health*, 15(4).

Chard, A.N., Trinies, V., Moss, D.M., Chang, H.H., Doumbia, S., Lammie, P.J. and Freeman, M.C. (2018) 'The impact of school water, sanitation, and hygiene improvements on infectious disease using serum antibody detection', *PLoS Neglected Tropical Diseases*, 12(4).

Charmaz, K. (2015) 'Grounded Theory: Methodology and Theory Construction', *International Encyclopedia of the Social and Behavioral Sciences (2nd Edition)*, pp. 402-407.

Chatterley, C., Slaymaker, T., Badloe, C., Nouvellon, A., Bain, R. and Johnston, R. (2018) 'Institutional WASH in the SDGs: data gaps and opportunities for national monitoring', *Journal of Water, Sanitation and Hygiene for Development*, 8(4), pp. 595-606.

Cohen, L., Manion, L. and Morrison, K. (2018) *Research Methods in Education (8th Edition)*. Abingdon: Taylor and Francis.

Corbin, J. and Strauss, A. (2015) *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (4th Edition)*. California: SAGE.

Creswell, J.W. and Miller, D.L. (2000) 'Determining validity in qualitative inquiry', *Theory into Practice*, 39(3), pp. 124-130.

Darteh, B., Moriarty, P. and Huston, A. (2019) *Building strong WASH systems for the SDGs (working paper)*. The Hague: IRC.

Das, S. (2020) 'Parched India - A looming crisis', *Journal of the Geological Society of India*, 95, pp. 333-336.

De Albuquerque, C. (2014) *Realising the Human Rights to Water and Sanitation: A Handbook (Introduction)*. Available at: https://reliefweb.int/report/world/realising-human-rights-water-and-sanitation-handbook (Accessed: 9 September 2019).

Dey, N.C., Parvez, M., Islam, M.R., Mistry, S.K and Levine, D.I. (2019) 'Effectiveness of a community-based water, sanitation, and hygiene (WASH) intervention in reduction of diarrhoea among under-five children: Evidence from a repeated cross-sectional study (2007–2015) in rural Bangladesh', *International Journal of Hygiene and Environmental Health*, 222(8), pp. 1098-1108.

Education World (2018) *Teachers spend only 19.1 per cent time teaching: NUEPA report*. Available at: https://www.educationworld.in/teacher

s-spend-only-19-1-per-cent-time-teaching-says-nuepa-report/ (Accessed: 18 September 2020).

Ejelonu, A., Feng, H. and McKeon, T. (2020) 'Evaluating WASH (Water, Sanitation and Hygiene) interventions in rural schools of West Bengal, India', *wH2O: The Journal of Gender and Water*, 7(1), pp. 17-25.

Ellis, A., McClintic, E.E., Awino, E.O., Caruso, B.A., Arriola, K.R., Ventura, S.G., Kowalski, A.J., Linabarger, M., Wodnik, B.K., Webb-Girard, A. and Muga, R. (2020). 'Practices and perspectives on latrine use, child feces disposal, and clean play environments in Western Kenya', *The American Journal of Tropical Medicine and Hygiene*, 102(5), pp. 1094-1103.

Engward, H. (2013) 'Understanding grounded theory', *Nursing Standard*, 28(7), pp. 37-41.

Eppig, C., Fincher, C.L. and Thornhill, R. (2010) 'Parasite prevalence and the worldwide distribution of cognitive ability', *Proceedings of the Royal Society B: Biological Sciences*, 277(1701), pp. 3801-3808.

Erismann, S., Diagbouga, S., Schindler, C., Odermatt, P., Knoblauch, A.M., Gerold, J., Leuenberger, A., Shrestha, A., Tarnagda, G., Utzinger, J. and Cissé, G. (2017) 'School children's intestinal parasite and nutritional status one year after complementary school garden, nutrition, water, sanitation, and hygiene interventions in Burkina Faso', *The American Journal of Tropical Medicine and Hygiene*, 97(3), pp. 904-913.

Fink, G., Günther, I. and Hill, K. (2011) 'The effect of water and sanitation on child health: evidence from the demographic and health surveys 1986–2007', *International Journal of Epidemiology*, 40(5), pp. 1196-1204.

Garn, J.V., Trinies, V., Toubkiss, J. and Freeman, M.C. (2017) 'The role of adherence on the impact of a school-based water, sanitation, and hygiene intervention in Mali', *The American Journal of Tropical Medicine and Hygiene*, 96(4), pp. 984-993.

Ghanim, M., Dash, N., Abdullah, B., Issa, H., Albarazi, R. and Al Saheli, Z. (2016) 'Knowledge and practice of personal hygiene among primary school students in Sharjah-UAE', *Journal of Health Science*, 6(5), pp. 67-73.

Glaser, B.G. (1992) *Emergence vs Forcing: Basics of Grounded Theory Analysis*. California: Sociology Press.

Government of NCT of Delhi (Planning Department) (2020) 'Education (Chapter 15)', in *Economic Survey of Delhi (2019-2020)*. Delhi: Government of NCT of Delhi.

Gupta, A., Khalid, N., Desphande, D., Hathi, P., Kapur, A., Srivastav, N., Vyas, S., Spears, D. and Coffey, D. (2019) *Changes in open defecation in rural north India: 2014-2018*. Bonn: Institute of Labor Economics.

Halai, N. (2007) 'Making use of bilingual interview data: Some experiences from the field', *The Qualitative Report*, 12(3), pp. 344-355.

Harahap, T.A., Saefuddin, A., Riani, E. and Indriyanto, B. (2018) 'The relationship between clean school environment and student's clean

lifestyle behaviour in Indonesia junior and senior high school', *International Journal of Scientific and Technology Research*, 7(3), pp. 100-104.

Hennegan, J., Kibira, S.P., Exum, N.G., Schwab, K.J., Makumbi, F.E. and Bukenya, J. (2020) ‘‘I do what a woman should do’: a grounded theory study of women’s menstrual experiences at work in Mukono District, Uganda’, *BMJ Global Health*, 5(11), pp. 1-11.

Hetherington, E., Eggers, M., Wamoyi, J., Hatfield, J., Manyama, M., Kutz, S. and Bastien, S. (2017) 'Participatory science and innovation for improved sanitation and hygiene: Process and outcome evaluation of Project SHINE, a school-based intervention in rural Tanzania', *BMC Public Health*, 17(1).

National Herald (2021) *New session of schools to begin with studies from home [Online]*. Available at: https://www.nationalheraldindia.com/national/new-session-of-schools-to-begin-with-studies-from-home (Accessed: 7 June 2021).

Hindustan Times (2021) *Delhi: Municipal corporation employees go on indefinite strike over non-payment of salaries [Online]*. Available at: https://www.hindustantimes.com/delhi-news/delhi-municipal-corporation-employees-go-on-indefinite-strike-over-non-payment-of-salaries/story-qJB6FAQGTQFaRm3pQFhLuI.html (Accessed: 1 March 2021).

Humphrey, J.H. (2009) 'Child undernutrition, tropical enteropathy, toilets, and handwashing', *The Lancet*, 274, pp. 1032-1035.

Humphrey, J.H. (2019) 'Reducing the user burden in WASH interventions for low-income countries', *The Lancet Global Health*, 7(9), pp. 1158-1159.

Jain, S., Khurana, N. and Bajaj, B. (2018) 'Involvement led attitudes towards responsible waste management under Swachh Bharat campaign: A study of government employees working in National Capital Region of India', *International Journal of Indian Culture and Business Management*, 16(4), pp. 384-399.

Johns Hopkins University (2021) *Coronavirus Resource Center: COVID-19 Dashboard*. Available at: https://coronavirus.jhu.edu/map.html (Accessed: 3 May 2021).

Joshi, D., Kooy, M. and Van den Ouden, V. (2016) 'Development for children, or children for development? Examining children's participation in School-Led Total Sanitation programmes', *Development and Change*, 47(5), pp. 1125-1145.

Kalra, A. (2019) *In elections, teachers are like Malko from 'Newton', supportive and unsung*. Available at: https://thewire.in/politics/elections-teachers-malko-newton (Accessed: 17 August 2020).

Karon, A.J., Cronin, A.A., Cronk, R. and Hendrawan, R. (2017) 'Improving water, sanitation, and hygiene in schools in Indonesia: A cross-sectional assessment on sustaining infrastructural and behavioral interventions', *International journal of Hygiene and Environmental Health*, 220(3), pp. 539-550.

Khalil, I.A., Troeger, C., Rao, P.C. Blacker, B.F., Brown, A., Brewer, T.G.

Colombara, D.V., De Hostos, E.L., Engmann, C. and Guerrant, R.L. (2018) 'Morbidity, mortality, and long-term consequences associated with diarrhoea from Cryptosporidium infection in children younger than 5 years: a meta-analyses study', *The Lancet Global health*, 6(7), pp. e758-e768.

La Con, G., Schilling, K., Harris, J., Person, B., Owuor, M., Ogange, L., Faith, S. and Quick, R. (2017) 'Evaluation of student handwashing practices during a school-based hygiene program in rural Western Kenya, 2007', *International Quarterly of Community Health Education*, 37(2), pp. 121-128.

Lewis, H.E., Greenland, K., Curtis, V. and Schmidt, W.P. (2018) 'Effect of a school-based hygiene behavior change campaign on handwashing with soap in Bihar, India: cluster-randomized trial', *The American Journal of Tropical Medicine and Hygiene*, 99(4), pp. 924-933.

Lincoln, Y.S. and Guba, E. (1985) *Naturalistic Inquiry*. California: SAGE.

Longhurst, R. (2016) 'Semi-structured Interviews and Focus Groups (Chapter 9)', in Clifford, N., Cope, M., Gillespie, T., French, S. (ed.) *Key Methods in Geography*. California: SAGE.

Malterud, K. (2001) 'Qualitative research: standards, challenges, and guidelines', *The Lancet*, 358(9280), pp. 483-488.

Mara, D. (2017) 'The elimination of open defecation and its adverse health effects: a moral imperative for governments and development professionals', *Journal of Water, Sanitation and Hygiene for Development*, 7(1), pp. 1-12.

Martin, N.A., Hulland, K.R.S., Dreibelbis, R., Sultana, F. and Winch, P.J. (2018) 'Sustained adoption of water, sanitation and hygiene interventions: Systematic review', *Tropical Medicine and International Health*, 23(2), pp. 122-135.

McGuinness, S.L., O'Toole, J., Forbes, A.B., Boving, T.B., Patil, K., D'Souza, F., Gaonkar, C.A., Giriyan, A., Barker, S.F., Cheng, A.C. and Sinclair, M. (2020a) 'A stepped wedge cluster-randomized trial assessing the impact of a riverbank filtration intervention to improve access to safe water on health in rural India', *The American Journal of Tropical Medicine and Hygiene*, 102(3), pp. 497-506.

McGuinness, S.L., O'Toole, J., Ayton, D., Giriyan, A., Gaonkar, C.A., Vhaval, R., Cheng, A.C. and Leder, K. (2020b) 'Barriers and enablers to intervention uptake and health reporting in a water intervention trial in rural India: a qualitative explanatory study', *The American Journal of Tropical Medicine and Hygiene*, 102(3), pp. 507-517.

McMichael, C. (2019) 'Water, sanitation and hygiene (WASH) in schools in low-income countries: A review of evidence of impact', *International Journal of Environmental Research and Public Health*, 16(359).

Michie, S., Van Stralen, M.M. and West, R. (2011) 'The behaviour change wheel: A new method for characterising and designing behaviour change interventions', *Implementation Science*, 6(42).

Mitul, M. and Agha, E. (2018) *Teachers spend only 19% of their time in educating students, poll duty and surveys take up rest*. Available at: https://www.news18.com/news/india/teachers-spend-only-19-of-their-time-in-educating-students-poll-duty-and-surveys-take-up-rest-1884135.html (Accessed: 17 August 2020).

Mohapatra, G. (2019) 'Projected behavioural change in Swachh Bharat Mission: a public policy perspective', *Indian Journal of Public Administration*, 65(2), pp. 451-474.

Mond, H. and Prakash, P. (2019) 'Motivations to set up and manage low-fee private schools in India', *Education Policy Analysis Archives*, 27(134).

Mulopo, C. and Chimbari, M.J. (2021) 'Water, sanitation, and hygiene for schistosomiasis prevention: a qualitative analysis of experiences of stakeholders in rural KwaZulu-Natal', *Journal of Water, Sanitation and Hygiene for Development*, 11(2), pp. 255-270.

Okello, E., Kapiga, S., Grosskurth, H., Makata, K., Mcharo, O., Kinungh'i, S. and Dreibelbis, R. (2019) 'Factors perceived to facilitate or hinder handwashing among primary students: a qualitative assessment of the Mikono Safi intervention schools in NW Tanzania' *BMJ Open*, 9(11).

Orgill-Meyer, J. and Pattanayak, S.K. (2020) 'Improved sanitation increases long-term cognitive test scores', *World Development*, 132.

Patton, M.Q. (1987) *How to Use Qualitative Methods in Evaluation*. California: SAGE.

Pinkerton, R., Oriá, R.B., Lima, A.A.M., Rogawski, E.T., Oriá, M.O.B., Patrick, P.D., Moore, S.R., Wiseman, B.L., Niehaus, M.D. and Guerrant, R.L. (2016) 'Early childhood diarrhea predicts cognitive delays in later childhood independently of malnutrition', *The American Journal of Tropical Medicine and Hygiene*, 95(5), pp. 1004-1010.

Price, J.H. and Murnan, J. (2004) 'Research limitations and the necessity of reporting them', *American Journal of Health Education*, 35(2), pp. 66-67.

Rahman, M.H.U., Malik, M.A., Chauhan, S., Patel, R., Singh, A. and Mittal, A. (2020) 'Examining the linkage between open defecation and child malnutrition in India', *Children and Youth Services Review*, 117.

Ray, I. (2020) 'Viewpoint - Handwashing and COVID-19: Simple, right there…?', *World Development*, 135.

Ritchie, H. and Roser, M. (2019) *Our World in Data - Sanitation*. Available at: https://ourworldindata.org/sanitation (Accessed: 14 December 2020).

Saboori, S., Greene, L.E., Moe, C.L., Freeman, M.C., Caruso, B.A., Akoko, D. and Rheingans, R.D. (2013) 'Impact of regular soap provision to primary schools on hand washing and E. coli hand contamination among pupils in Nyanza Province, Kenya: a cluster-randomized trial', *The American Journal of Tropical Medicine and Hygiene*, 89(4), pp. 698-708.

Sandberg, E.H. and McCullough, M.B. (2010) 'The Development of Reasoning Skills (Chapter 10)' in Sandberg, E.H. and Spritz, B.L. (ed.) *A*

Clinician's Guide to Normal Cognitive Development in Childhood. New York: Routledge.

Scott, K.W. (2004) 'Relating categories in grounded theory analysis: Using a conditional relationship guide and reflective coding matrix', *The Qualitative Report*, 9(1), pp. 113-126.

Shrestha, A., Schindler, C., Odermatt, P., Gerold, J., Erismann, S., Sharma, S., Koju, R., Utzinger, J. and Cissé, G. (2020) 'Nutritional and health status of children 15 months after integrated school garden, nutrition, and water, sanitation and hygiene interventions: a cluster-randomised controlled trial in Nepal', *BMC Public Health*, 20(158).

Staniford, L.J. and Schmidtke, K.A. (2020) 'A systematic review of hand-hygiene and environmental-disinfection interventions in settings with children', *BMC Public Health*, 20(195).

Troeger, C., Forouzanfar, M., Rao, P.C., Khalil, I. Brown, A., Reiner Jr, R.C., Fullman, N., Thompson, R.L., Abajobir, A. and Ahmed, M. (2017) 'Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015', *The Lancet Infectious Diseases*, 17(9), pp. 909-948.

United Nations Children's Fund (UNICEF) (2020) *India - Key Demographic Indicators (Multiple Indicator Cluster Survey)*. Available at: www.data.unicef.org/country/ind (Accessed: 7 May 2020).

United Nations General Assembly (2015) *Transforming Our World: The 2030 Agenda for Sustainable Development*. New York: Assembly, U.N.G.

Wagner, J.T. and Samuelsson, I.P. (2019) 'WASH from the START: Water, sanitation and hygiene education in preschool', *International Journal of Early Childhood*, 51(1), pp. 5-21.

Wong, L.P. (2008) 'Data analysis in qualitative research: A brief guide to using NVivo', *Malaysian Family Physician*, 3(1), pp. 14-20.

World Bank (2020) *Prevalence of stunting, height for age (% of children under 5) - India*. Available at: https://data.worldbank.org/ (Accessed: 12 October 2020).

World Health Organisation (WHO) (2015) *Stunting in a nutshell*. Available at: https://www.who.int/nutrition/healthygrowthproj_stunted_videos/en/ (Accessed: 5 June 2019).

Appendix 1. School staff interview Questions (principal, teachers, cleaners)

Thank you very much for taking the time to speak to us. Just a reminder, we are representatives of Newcastle University conducting an independent evaluation of the work that was done in your school by [NGO], and we want to hear your views and opinions on the impact it has had on the school. So please feel free to speak your opinions to us.

1. Please tell us a bit more about your role at the school. What do you do day to day?

2. How long have you worked at the school for?

3. Can you explain to us a bit more about the work [NGO] has done in the school?

4. Do you think [NGO's] work has helped to bring solutions to problems faced by children and staff at your school?

5. What sort of impact do you think that [NGO's] work has had on the school? How come?

6. Is there anything about [NGO's] work that you would change or want to improve?

7. What would you say is the most significant change that has happened in the school as a result of [NGO's] work?

8. What would you say is the most significant change for the school children?

9. Do you think that [NGO's] work in this school has had any impact on the local community? Could you explain this please?

What are the costs and benefits of Teach First's participant placement strategy at individual, school and policy levels?: A progress report

Sally J McDonald

ABSTRACT: Through my ongoing doctoral education study (EdD) I am aiming to explore the costs and benefits of Teach First's participant placement strategy at individual, school, and policy levels. In this progress report I will provide an overview of the background and rationale for the study, detail the methodology, outline the progress made and complications faced to date, and reflect on the findings so far.

Key words: *teacher training, placement, retention, Teach First*

Contact: S.J.McDonald2@newcastle.ac.uk

Background and Rationale

What is Teach First?

Teach First is an educational charity which aims to end educational inequality and build a fair education through its three strands: training teachers, developing leaders, and building school networks (Teach First, 2021a). This study is focused on the teacher training programme. The programme begins with an intensive five-week course known as Summer Institute, after which participants spend most of their time teaching 60-80% of a full timetable in schools, supported by in-school mentors and university tutors, with a few additional training days and conferences throughout the school year (Teach First, 2021b). The Teach First programme differs from other routes into teaching because participants are paid a salary from the outset and there is a greater emphasis on progression into leadership. Furthermore, the charity only partners with schools where a high percentage of pupils come from low-income backgrounds (identified through the Income Deprivation Affecting Children Index) (Teach First, 2021c) and thus supports the national policy aim of ensuring 'Educational Excellence Everywhere' (DfE, 2016).

Teach First's placement strategy

As a result of the charity's aim to work with these types of schools, participants are sometimes required to relocate in order to complete the programme; they are given the opportunity to rank their location preferences, but ultimately the needs of schools and pupils are prioritised in the placement strategy so not all receive their first choice (Teach First, 2021b). Furthermore, some of the regions cover large geographical areas. The North East, for example, stretches from the Scottish border to North Yorkshire (Teach First, 2021d). Consequently, even if participants

receive their first choice of region, they may still end up placed in a school which is quite far from where they wanted to be.

Potential issues with this placement strategy

This strategy is similar to teacher placement strategies used elsewhere, such as parts of the USA (Teach for America, 2019) and France (Cole, 2001). However, as will be outlined, issues relating to teacher wellbeing, recruitment and retention are prominent in the UK, raising questions about the impact of the strategy at individual, school and national levels.

Concerns at an individual level

Teaching has been found to be a stressful occupation (Bachkirova, 2005; Jepson and Forrest, 2006). Labour Force Surveys from 2013-2016 found that work-related stress, anxiety and depression were 45% higher in the education sector than the average across all industry groups (HSE, 2016) and the Workload Challenge launched by the Department for Education in 2015 acknowledged the intense workload associated with teaching (DfE, 2015). Teach First participants seem especially vulnerable to wellbeing challenges given that they must teach an 80% timetable whilst simultaneously meeting the demands of other teaching routes (Teach First, 2014), often in particularly challenging schools (Allen et al, 2016). This raises the question of how individual participants' wellbeing is impacted when, on top of these challenges, they do not receive their first choice of region. Existing literature on Teach First (Muijs et al, 2010; Allen & Allnutt, 2013; Allen et al, 2016) tends to focus primarily on the impact it has on schools and students. However, it does not consider how it affects participants themselves. I therefore conducted a small-scale Masters study into participants' experiences and found that, at least in some instances, there was a link between placement location, wellbeing and retention. Some of the participants who did not get their first choice of region felt that it compounded other challenges and impacted negatively on their experiences. Although this was only a small study, these findings suggest that this is an area worthy of further study.

School level concerns

There is a consensus that high teacher turnover correlates negatively with student achievement (Guin, 2004; Boyd et al, 20011; Hanushek et al, 2005; Ronfeldt et al, 2013) and impacts negatively on staff cohesion and finances (Byrk & Schneider, 2002; Barnes et al, 2007), and the schools eligible to partner with Teach First typically find recruitment and retention to be especially challenging (Allen & Allnutt, 2013). Furthermore, studies which have noted the positive impact of Teach First on school performance (Muijs et al, 2010; Allen & Allnutt, 2013) did not identify a positive impact until the participants' second year of teaching, meaning that a partner school only really benefits for one year if participants choose to leave on completion of the programme. Teach

First is only a two-year programme so it is always possible that participants will leave after that two-year commitment, but it seems more likely that they will leave if they did not want to be in that area in the first place. This raises the question of whether it is worth the short-term benefits if it increases the long-term problems associated with high teacher turnover for schools?

The national picture

The falling retention rate within the teaching profession has become a key concern in recent years (Clayes, 2017; Foster, 2019), and retention problems tend to be most acute in schools in disadvantaged areas like those eligible for Teach First (DfE, 2016). Furthermore, the average cost of training a teacher through Teach First is £14,000 more than any other teacher training route (Allen et al, 2016), so the overall benefits of the programme are arguably highly contingent on the number of participants who remain in teaching. This raises a potential issue with nationwide impact of the strategy. If participants have negative experiences as a result of their placement schools or locations, and choose to leave the profession as a result, how does this impact upon the profession as a whole?

Research Questions

The study therefore aims to answer the following research questions:

1. How do Teach First participants placed in the North East of England feel that Teach First's placement strategy impacts on their experiences during their two years on the programme?

2. How do staff in Teach First partner schools in the North East feel that the participant placement strategy impacts on participants and schools?

3. What is the relationship between Teach First's participant placement strategy and retention on and beyond the two-year programme?

Methodology

The conceptual framework shown in figure 1 shows how these research questions are interlinked. The micro level impacts on participants (explored by research question 1) influence the meso level impacts on schools (explored by research question 2); despite individuals' intentions when beginning the Teach First programme, a positive placement experience may encourage them to remain in their schools after completing the programme, whereas a negative experience may sway them to leave, decreasing the long-term benefits for the partner schools. Furthermore, retention at a school level feeds into the macro level of national teacher retention trends: research question 3 explores how the benefits of getting teachers into schools which struggle to recruit teachers in the short-term has to be balanced with the need for teacher training to be a positive experience in order to improve retention in the profession nationally.

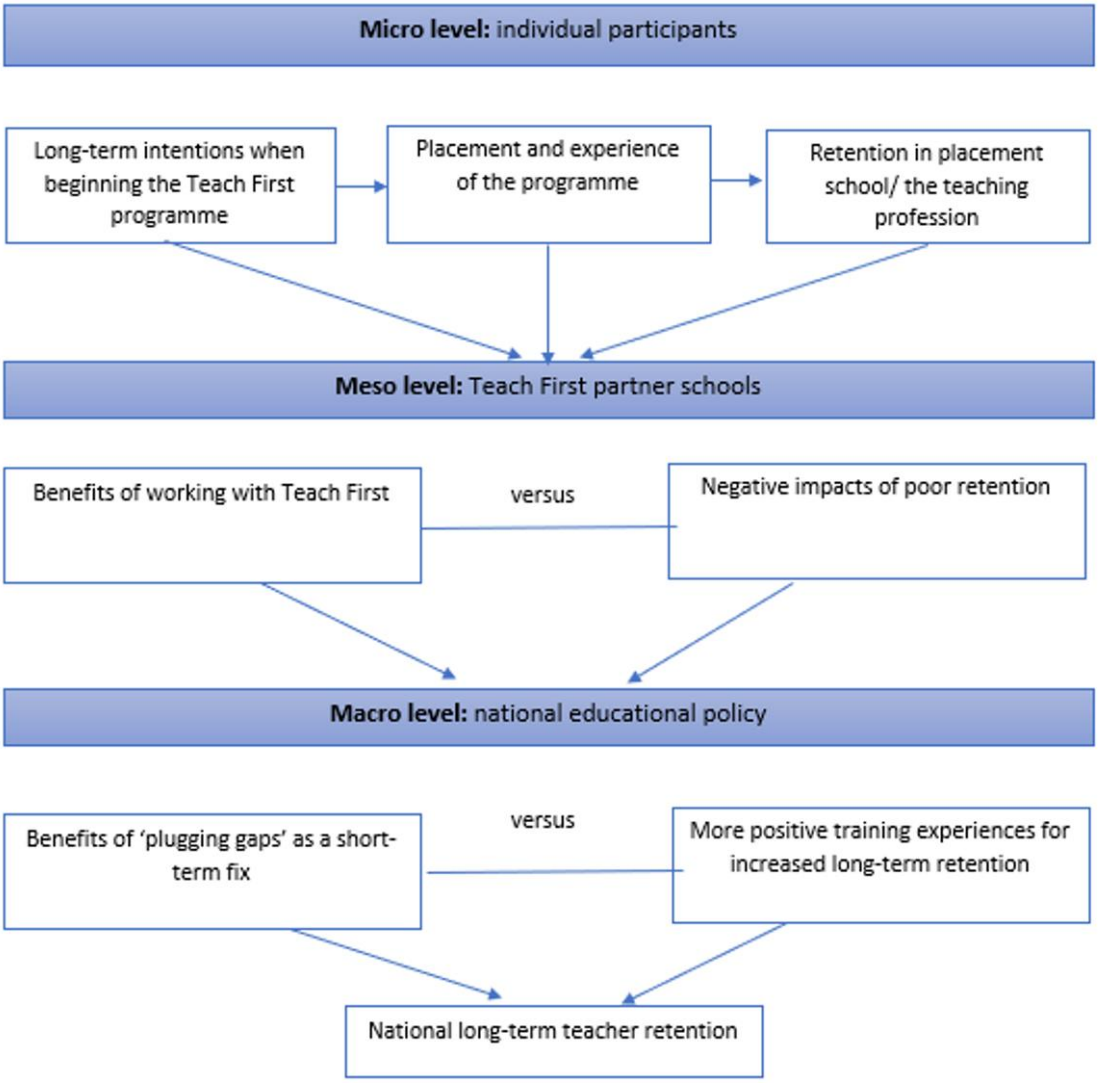


Figure 1. Conceptual framework

In order to answer research question 1, semi-structured interviews are being conducted with participants in the North East region at different stages of the programme. Members of the 2019 cohort were interviewed near the end of their first year and will be interviewed again once they have almost completed the programme, while those from the 2020 cohort were interviewed just before starting the programme and will be interviewed again towards the end of their first year. Due to the timing of the study, members of the 2018 cohort were interviewed just once, towards the end of their time on the programme. Interviews were favoured over surveys as it was thought that teachers' high workload (DfE, 2015) may deter them from completing detailed qualitative surveys. They were also chosen over focus groups because participants were to be asked about the challenges

they had faced on the programme and it was thought that they would be more comfortable discussing any struggles they had faced privately.

In order to answer research question 2, interviews are being conducted with staff in Teach First partner schools who either work closely with participants (for example as mentors, or those involved in leading teacher training within school) or those who are involved in working with Teach First to arrange participant placements within their schools (typically members of the Senior Leadership Team). Unlike the participants, who will be reinterviewed, the school staff will just be interviewed once on the basis that they are more abstracted from the programme and therefore their responses are less likely to noticeably shift over the course of a year than the participants, who are on a steep learning curve and more emotionally invested in the programme, and therefore are arguably more likely to change their views over the course of time.

Finally, in order to answer research question 3, a sizeable dataset has been obtained from Teach First's participant database. This anonymised data includes the cohort year, recruitment year, first choice of region and allocated region for each participant, as well as whether or not they completed the two-year programme. For most, it also states whether or not they stayed in their placement school or the local area the following year, although there are some gaps here as this was reliant on participants self-reporting their destinations. Due to changes in Teach First's data consent arrangements, the dataset only includes those who began the programme between 2015 and 2017. However, this is still a sizeable dataset (with approximately 4,500 entries), so will be very helpful when it comes to analysing national trends related to placement location, retention within the area and retention in the profession as a whole.

The study is aligned to the constructivist paradigm, where knowledge is viewed as an evolving human construct, influenced by the status, experiences and values of respondents and researchers (Pring, 2000), as it seeks a rich understanding of participants' experiences of the programme, which are deeply personal. However, a pragmatic, mixed methods approach is being taken because while the interviews will offer qualitative insights into the lived experiences of participants and school staff, revealing some of the personal realities behind the national retention trends, the quantitative analysis will broaden the scope of the study and enable comparison of the qualitative findings with national trends. In line with BERA (2018), ethical concerns will be paramount throughout the study with informed consent gained and participants granted the right to withdraw at any time. See figure 2 for the timescale of the research.

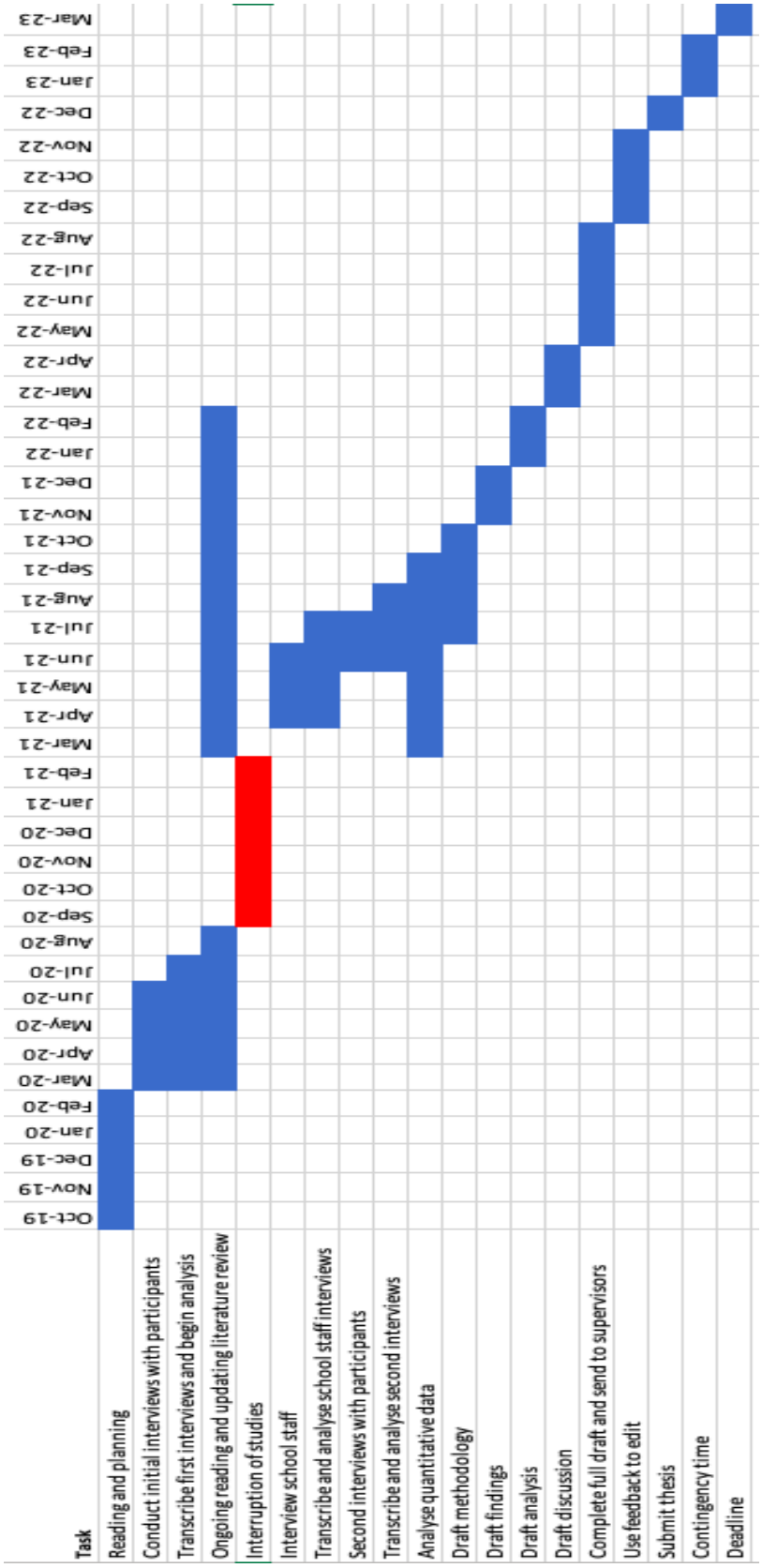


Figure 2. Gantt chart of research project timescale.

Progress to date

So far, I have conducted and transcribed twelve semi-structured interviews with Teach First participants, in relation to research question 1. Of these, three were in their final year of the programme, three were in their first year and six were just about to begin their Summer Institute training. Seven of these participants had selected the North East as their first choice, three stated that they did not have a location preference, and two had selected somewhere other than the North East. Furthermore, although they were all teaching in the North East region, their placement schools spanned from Newcastle to Redcar, so they were able to comment on the impact of their varied placement locations within the broader Teach First region. Those who were in their first year and those who were about the begin the programme will be interviewed again at the end of this academic year to see if and how their views have changed over the course of a year.

I am currently in the process of arranging and conducting interviews with relevant staff in Teach First partner schools in order to answer research question 2. At the time of writing, I have interviewed three senior school leaders, all of whom are involved in liaising with Teach First to get participants into their schools, then supporting them and overseeing their training once they are there. Finally, in order to answer research question 3, I have begun to explore the dataset from Teach First, considering the proportion of participants who got their first choice of region, how this links to completion rates and retention in schools and areas following the two-year commitment, and how this all varies between regions.

Preliminary findings and discussion

Although it is too early to draw any conclusions, the interviews conducted so far indicate that, at least in some cases, Teach First's placement strategy is a contentious issue for both participants and schools, and early analysis of the quantitative data shows a connection between placement and retention.

In the interviews with participants, the impact of the intense workload of the Teach First programme on wellbeing emerged as a strong theme, emphasising the need for the placement strategy to support wellbeing wherever possible. Furthermore, it was clear from the interviews that the level of support participants receive from their placement schools varies widely. Arguably this is something that could be taken into account, considering whether the negative impacts of not receiving a first choice of region can be somewhat mitigated by at least being placed a more supportive school, although this may be hard to measure. Participants' views about being placed in the North East were mixed. Those who had put the North East as their first choice all felt that getting their preferred area had been important in helping them due to their existing support networks. Some participants were flexible and open-minded about location as they had no ties to any particular area. Those in

that position in the 2018 and 2019 cohorts threw themselves into the Teach First experience, building a tight network with their fellow North East participants. Those in the 2020 cohort were worried about how they were going to do this during the pandemic. However, this approach was not for everyone. One of the 2019 cohort lamented that as they had not received their first choice of region, they were struggling to meet anyone outside of education and therefore found it difficult to switch off from work.

Furthermore, the interviews indicate that placement area within the larger North East region plays a key role in shaping participants' experiences. Generally, those who were placed within big cities seemed to be enjoying the experience more, whereas those in more isolated areas were struggling. One participant who had not received their first choice of region was having a particularly negative experience, so much so that they were considering leaving the profession. They had been placed in a remote area of the region and felt that this had compounded the challenges of relocation as they were further from transport links (particularly limiting as they could not drive), more removed from most of their fellow participants, and further from the partner university, making societies (a possible source of new friendships) inaccessible. Again, this suggests that perhaps the negative impacts of a participant not receiving their first choice of region could be partly mitigated if they were placed in their preferred area within their allocated region.

In the interviews with school staff, most were very positive about Teach First; they felt that partnership supports recruitment and provides them with high quality teachers, some of whom do stay beyond two years. However, one had a more cynical view and seemed to view Teach First as a conveyer belt. They often struggled to recruit and found that Teach First helped with this, but they then felt that participants were not always invested in the school because they had no intention of staying in the area beyond the two-year commitment, and several of their participants had dropped out. However, they kept returning to Teach First for more participants again, because it solved their recruitment difficulties.

When reflecting on their participants' experiences, even the staff who had spoken positively of Teach First noted the challenges faced by participants who did not get their first choice of region, at least at first. One compared two participants who they had mentored, noting that one was able to go home to a supportive family and a cooked meal each night, while the other knew no one in the area and had therefore required additional support from the school at first. They did state that the latter eventually found their feet and settled well, but this nonetheless highlights the challenges that the strategy raises.

Initial study of the national quantitative data has revealed that 71% of participants receive their first choice of region whereas 26% do not (a further 3% were flexible and did not select a first choice). However, there are marked differences in the proportion of

participants in each region who received their first choice; 81% of those placed in the West Midlands had wanted to be placed there compared to just 46% of those who were placed in the South West, for example. Interestingly, it does not seem that receiving their first choice of region affects participants' retention on the programme. In fact, the completion rate is actually slightly higher for those who do not get their first choice (88% compared to 87%). However, a negative effect is shown after two years when just 19% of those who do not receive their first choice of region remain in their school compared to 29% of those who do, and just 35% of those who do not get their first choice remain teaching in their area compared to 56% of those who do. Furthermore, in the data about participants' destinations following completion of the programme, which was self-reported to Teach First at the end of the two years, there were more gaps in the records of those who had not received their first choice, possibly suggesting a lower rate of engagement with Teach First in general. Interestingly, the data also revealed that there is an overlap each year, where some participants do not get their first choice of region, but others are put there against their preference, and therefore the matching can be described as 'unstable' (Gale & Shapley, 1962). The reasons for this unstable matching require further exploration, as this reveals that there is more to the process than some regions being oversubscribed.

Changes to planned data collection and its implications on findings

Unfortunately, the Covid-19 pandemic has impacted this study in several ways. Firstly, interviews began in person within schools but when the first lockdown started, the remaining interviews had to be conducted online. This created an undesirable but unavoidable lack of consistency. This is because interview context has been found to have a profound impact on data (Clarke & Robertson, 2001; Denzin & Lincoln, 2005). As Mann (2016) argues, interviewing in an interviewees home may offer access to different perspectives than interviewing in the workplace, as employees may be influenced by their emotional ties to their institutional setting. Although Mann was referring to in-person interviews in the home, this is still relevant because those who were interviewed virtually from home were still afforded with a sense of abstraction from their school in comparison to those who were interviewed in person in schools.

Furthermore, current participants' experiences of Teach First will inevitably have differed significantly from those of previous cohorts because schools have been closed to all but the children of key workers for several months over the past year, with the majority of children learning online. Additionally, the 2020 Summer Institute all took place online and they have had less in-person training and support throughout the year. Even outside of school life, lockdowns may have impacted upon the ways in which they could experience their placement area, with 'Stay at home' and 'Stay

local' orders and many local amenities closed (Cabinet Office, 2021).

The full impact of the pandemic will not become clear until the next set of interviews are conducted. I hypothesise that this is likely to impact upon participants' views about their placement. For some, not having the opportunity to get to know their local area or network in person may have made settling in harder. However, due to a greater emphasis on virtual spaces, this may have made physical location less important than in previous years. There may be other effects of the pandemic also, with interviewees who completed the programme in 2020 stating how it has influenced their career paths. For example, one of the final year participants had chosen to remain in their placement school for a third year, but only because their plan to teach abroad had fallen through due to covid.

Conclusion

This mixed methods study will provide insights into the individual, school and policy level effects of Teach First's participant placement strategy. A range of perspectives will be considered including how participants' thoughts change over the duration of their training experience, something that is pertinent in light of current issues in teacher recruitment, retention and wellbeing. Ultimately, it is hoped that this study will support Teach First, as evaluating this aspect of its strategy could eventually lead to strategic improvements and increase its impact. Although this study is focused primarily on the North East, respondents include a range of participants, including school staff, some participants who got their first choice of region and some who did not, and comparison with quantitative national data. This will broaden the scope and enable greater comparisons. Consequently, it should provide useful data on an area which is absent from existing literature. Furthermore, although not the original plan, due to the timing of the research, it also seems likely that it will contribute greatly to the current educational zeitgeist, offering insights into what it was like to train as a teacher during the Covid-19 pandemic.

References

Allen, R. and Allnutt, J. (2013) *Matched panel data estimates of the impact of Teach First on school and departmental performance*, London, Institute of Education.

Allen, R., Belfield, B., Greaves, E., Sharp, C. and Walker, M. (2016) *The Longer-Term Costs and Benefits of Different Initial Teacher Training Routes,* London, The Institute of Fiscal Studies.

Bachkirova, T. (2005) *Teacher stress and personal values: an exploratory study*, School Psychology International, (Vol.26, No.3), pp.340-352.

Barnes, G., Crowe, E. and Schaefer, B. (2007) *The costs of turnover in five districts. A pilot study,* Washington DC, National Commission on Teaching and America's Future.

British Educational Research Association. (2018) *Ethical Guidelines*

for Educational Research (fourth edition), London, BERA.

Boyd, D., Lankford, H., Loeb, S., Ronfeldt, M. and Wyckoff, J. (2011) *The role of teacher quality in retention and hiring: Using applications-to-transfer to uncover preferences of teachers and schools*, Journal of Policy and Management, (Vol.30, No.10), pp.88–110.

Bryk, A.S., and Schneider, B. (2002) *Trust in schools: A core resource for improvement*, New York, Russell Sage Foundation.

Cabinet Office. (2021) *National Lockdown: Stay at Home* [online] Available at https://www.gov.uk/guidance/national-lockdown-stay-at-home#national-lockdown-stay-at-home [accessed 12/3/2021].

Clarke, A., and Robertson, A. (2001) *Lifting a corner of the research rug: A case for meta-interviews in qualitative studies*.,Teaching and Teacher Education, (Vol.17, No.7), pp773–782.

Clayes, Z.D. (2017) *Teacher Retention: Are England's teachers leaving?* UK, NFER.

Cole, A. (2001) *The New Governance of French Education*?, Public Administration, (Vol.29 No.3), pp707-724.

Denzin, N.K. and Lincoln, Y.S. (Eds) (2005) *Handbook of qualitative research*, Sage, London.

Department for Education. (2015) *Government response to the Workload Challenge*, London, DfE.

Department for Education. (2016) *Educational Excellence Everywhere*, London, DfE .

Foster, D. (2019) *Teacher recruitment and retention in England*, UK, House of Commons Library.

Gale, D. and Shapley, L.S. (1962) *College Admissions and the Stability of Marriage,* The American Mathematical Monthly, (Vol.69, No.1), pp.9-15.

Guin, K. (2004) *Chronic teacher turnover in urban elementary schools*, Educational Evaluation and Policy Analysis, (Vol.12, No.42), pp1–25.

Hanushek, E. A., Kain, J.F., O'Brien, D.M., and Rivkin, S.G. (2005) *The market for teacher quality*, Cambridge, NBER.

Health and Safety Executive. (2016) *Work related Stress, Anxiety and Depression Statistics in Great Britain 2016.* UK, Office for National Statistics.

Jepson, E. and Forrest, S. (2006) *Individual contributory factors in teacher stress: the role of achievement striving and occupational commitment*, British Journal of Educational Psychology, (Vol.76, No.1), pp183-197.

Mann. S. (2016) *The Research Interview: Reflective Practice and Reflexivity in Research Processes,* UK, Palgrave Macmillan.

Mujis D., Chapman C., Collins, A. and Armstrong, P. (2010) *Maximum Impact Evaluation. The impact of Teach First teachers in schools: An evaluation*, London, Teach First.

Pring, R. (2000) *The 'False Dualism' of Educational Research*, Journal of

Philosophy of Education, (Vol.34, No.2), pp.247-260.

Ronfeldt, M., Loeb, S. and Wyckoff, J. (2013) *How Teacher Turnover Harms Student Achievement*, American Educational Research Journal, *(*Vol.50, No.1), pp4-36.

Teach First. (2014) *Programme Guide.* London, Teach First.

Teach First. (2021a) *Our Mission.* [online]. Available at: https://www.teachfirst.org.uk/our-mission [accessed 11/3/2021].

Teach First. (2021b) *Training Programme* [online]. Available at: https://www.teachfirst.org.uk/training-programme [accessed 11/3/2021].

Teach First. (2021c) *Potential in every child? Recruiting quality new teachers for your school* [online]. Available at: https://www.teachfirst.org.uk/sites/default/files/2021-02/1068%20Secondary%20schools%20flyer_2021.pdf [accessed 16/4/2021].

Teach First. (2021d) *Where you could teach* [online]. Available at: https://www.teachfirst.org.uk/knowledge-base/about-the-programme/where-you-could-teach [accessed 11/3/2021].

Teach For America (2019) *Placement* [online]. Available at: https://www.teachforamerica.org/how-to-join/placement [accessed 13/7/2019].

A longitudinal study to explore the social grouping of international students in Higher Education in the UK

Hanh Pho

ABSTRACT: This longitudinal qualitative research examined different forms of the social grouping of international students enrolling for taught postgraduate degrees in a university in the United Kingdom. Semi-structured interviews were conducted at the start (n=23) and at the end of the academic year (n=19). The study finds that social grouping on an 'internationalised' campus occurs early in the sojourn and once formed, the pattern remains unchanged until the end of the year. Lack of intercultural communication skills and lack of confidence in English skills are a few of the contributory factors. The students tend to hold an essentialist view of cultures and are reluctant to initiate and maintain contact with students perceived to be of the 'other' cultural group. Some implications for higher education institutions and educators are discussed to enhance social interactions among the students.

Key words: *intercultural communication, internationalisation of higher education, international students, social segregation, sojourn, social interactions, social essentialism*

Contact: hanhpho93@gmail.com

Introduction

The number of international students worldwide has increased exponentially over the past two decades, from nearly a million in 2002 to over 5.3 million in 2017 (Forest & Altbach, 2006; UNESCO, 2019). The United Kingdom (UK) is the country with the second largest population of international students (Gil, 2014; Tickle, 2014). Due to the emerging importance of this group of students, internationalisation has become a key agenda in UK HE (Nashrawan & Cox, 2016).

HE institutions in the UK have usually categorised students based on the fee status: (1) 'home' students (i.e., UK students), (2) EU students and (3) international students (i.e., students from other countries). However, in this study, the term 'international students' was used to refer to any non-UK students who temporarily relocate to a new country for an academic purpose (Pitts, 2009).

There is generally a lack of consensus on the meaning of internationalisation (Ray & Solem, 2009; Turner & Robson, 2007). For the purpose of this study, I employed the frequently cited work by Jane Knight (2004, p. 11) which defines 'internationalisation' as 'the process of integrating an international, intercultural or global dimension into the purpose, functions and delivery of post-secondary education'. Internationalisation is driven by the hope to strengthen mutual understanding by international mobility (Khoo, 2011); is expected to improve the quality of teaching and learning (de Wit, Hunter, Egron-Polak & Howard, 2015; Khoo, 2011) and enhance international connections and intercultural competence of the students (Brown, 2009c; Forsey, Broomhall, & Davis, 2012; Knight, 2011; Rizvi, 2011).

However, in many HE institutions in the UK, the USA and Australia, are at risk of exploitation (Kelly, 2000; Teichler, 2010) and have been regarded as 'cash cow' (CAUT, 2016). 'Internationalisation' has become a marketable element for many universities and has been driven by profit-seeking policies (DFATD, 2014; Warwick & Moogan, 2013); yet the extent to which these universities are truly 'international' remains questionable (Schartner & Cho, 2017).

The exposure to the international environment (e.g., studying with other students of various nationalities) and the international programming (e.g., participating in the international coursework or co-curricular activities) may not always result in meaningful communication (Forsey et al., 2012; Peacock & Harrison, 2009; Soria & Troisi, 2014). Many studies have repeatedly shown that interactions between international students and 'home' students of the host country are limited (Brown, 2009a, 2009b; Pho & Schartner, 2019; Schartner, 2015; Wright & Schartner, 2013). Social segregation, racial segregation and racism are not uncommon on 'internationalised' campus (Fincher & Shaw, 2009; Fincher & Shaw, 2011; Harper & Hurtado, 2007; Harper & Quaye, 2007; Kwon, Hernandez & Moga, 2019).

Research interest in social grouping in HE has been threefold: (1) studies focusing on social grouping with the academic environment (e.g., classroom interactions), (2) outside (i.e., accommodation halls and social events) and (3) on both occasions. This study focused on examining social grouping on both within and outside the academic environment.

Region of origin has been suggested to be one of the predictors on both contexts (Glass, Gómez & Urzua, 2014; Rienties & Nolan, 2014). Students coming from a country with a large cultural distance with the host culture usually struggle in socialising, hence, are more likely to be segregated from 'home' students (Gareis, 2012; Hanassab, 2006; Wang, 2016). International students who are readily accepted into the network of British students are more likely to originate from 'European or Anglophone origin' whose cultures are less distant (Harrison & Peacock, 2010, p. 884; Wang, 2016). A cross-sectional quantitative study by Glass, Gómez & Urzua (2014) shows that students from East, Southeast Asia and South Asia report significantly higher amount of contact with co-nationals and less contact with 'home' students than European students.

A considerable body of the literature highlights that the academic discourse in most 'Western' countries are characterised by critical thinking and logical argument (Hall, 1976; Hall, Held, & McGrew, 1992; Paul, 1982, 1993), whereas in East Asian countries, more emphasis is put on harmony of relationships and restraint of expressing personal ideas (Atkinson, 1997; Biggs, 1996). Students from the latter group are often portrayed in the literature as passive, less engaged with in-class participation and interactions (Harrison & Peacock, 2009, 2010).

However, this tendency to automatically 'revert' to cultural

differences and cultural deficits has recently been criticised by many scholars (Floyd, 2011; Hampden-Turner & Trompenaars, 2000; Nisbett et al., 2001). Nisbett et al. (2001) emphasise that despite some differences in cognitive processes, no empirical evidence of deficits in critical thinking amongst East Asian students has been found.

Floyd (2011) suggests that the heavy reliance on reproductive learning and the lack of in-class participation can be the result of language proficiency, rather than a cultural phenomenon. Indeed, low English proficiency has been reported in some studies as the main barrier in socialising with 'home' students (Montgomery & McDowell, 2009; Rientieset al., 2012) and affecting in-class interactions with other international colleagues (Peacock & Harrison, 2009).

In addition, 'home' students also express reluctance in instigating intercultural communication in the academic environment (e.g., multicultural group projects and in-class activities) (De Vita, 2002; Montgomery & McDowell, 2009; Rienties et al., 2012; Volet & Ang, 2012). For instance, a qualitative study by Harrison and Peacock (2010, p. 884) finds 'semi-distinct social spaces within the university environment' in which international and 'home' students rarely interact since the former is 'perceived as a threat' to the academic achievement due to their 'lack of understanding of British pedagogy'.

Some studies, meanwhile, suggest that intercultural interactions are limited partly because 'home' students have already formed their well-established network of friendships (Hendrickson, Rosen & Aune, 2011; Rienties et al., 2012); therefore, feel less inclined to build relationships with international students.

In general, most studies in the internationalisation of HE field have usually adopted an institutionalist perspective in which the culture, the educational and learning style of the host institution is seen as the 'standard' which international students need to adapt to. International students are often portrayed as a counterbalance (or a 'rival' group) to 'home' (or 'host national') students; and both said groups are treated in many studies as homogenous groups (see Brown, 2009b; Fincher & Shaw, 2011; Glass & Westmont, 2014; Harrison & Peacock, 2010; Peacock & Harrison, 2009). Yet, the former group is presumed to be the ones in need of 'adjusting' to form more social interactions with their 'host national' peers. Although regarded as a relatively more heterogenous cohort in some studies, nationality, ethnicity and cultures of the students are often used as categories of analysis (Durkin, 2008; Gareis, 2012; Glass, Gómez & Urzua, 2014; Wu & Hammond, 2011).

The institutionalist perspective is also in danger of elevating the role of 'home' students in the life of international students whilst ignores other 'noninstitutional' forms of interactions, such as contact among international students. It is generally unknown if other forms of the social grouping in HE may exist. Little effort has been made to monitor patterns of social grouping over time using

qualitative approach (Slaten et al., 2016).

This study, therefore, explored forms of the social grouping of international students in both academic and social settings. This study also aimed to discover rationales behind social grouping in UK HE. Within this context, primary research questions were:

1. What are the forms of social grouping that international students experience in the HE context in the UK?

2. What are the perceived factors influencing social grouping in the HE context in the UK?

Method

Longitudinal qualitative research design was applied, using semi-structured interviews, to monitor social grouping and to detect changes (if any) in the grouping pattern over time (De Vaus, 2001; Malhotra, 2010).

To maximise homogeneity across the sample, students enrolled in one-year taught master's programmes at a British university were recruited. These programmes were chosen because their large international student numbers and their similarity in structure and assessment. All students had an overall IELTS score of 6.5 and above or equivalent (the minimum language requirements for most postgraduate programmes in the UK) (Pho & Schartner, 2019; Young & Schartner, 2014).

Data were collected over nine months, from October – November 2017 to May – June 2018, referred in this paper as Phase 1 (P1) and Phase 2 (P2) respectively. Firstly, students were informed about the research at the end of the lecturers and were invited to participate in the research. 23 students participated in P1 and 19 students in P2 (a retention rate of 82.6%). Participation was voluntary and written consent was obtained prior to data collection.

Each interview was conducted for around 30 to 45 minutes on campus. Most were conducted in English (except one where the interviewee required to use her native language which was also the same as the researcher's). All interviews were audio recorded, transcribed verbatim and fully anonymised. More than 60% of respondents were female, aged around 20 – 30, and from East Asian and Southeast Asian countries, which broadly reflected the demographics of the larger student cohort in UK taught postgraduate degrees (Wright & Schartner, 2013; Schartner & Young, 2016).

NVivo, a computer software, was employed to manage the data. A total of 42 transcripts were generated from the interviews at two phases. Thematic analysis was applied, based on the work of Clarke and Braun (2014). First, I read transcripts a few times to 'familiarise' with the data. Then, key words and phrases were highlighted. Similarities were noted. These words or phrased were grouped together under the same category and labelled with a term (referred to as 'a code') (Clarke & Braun, 2014). Similar codes were put together and a term (known as a 'theme') to generalise their

meaning was developed (Clarke & Braun, 2014).

Criteria for evaluating this qualitative research

The interviewees communicated in English fluently, but sometimes, miscommunication might happen (e.g., using the same English words or terms but referring to different meanings). However, I minimised misunderstanding and maintain the credibility of interview data by employing the peer debriefing technique to cross-check the opinions of interviewees (Guba and Lincoln, 1989). During the second interview, I also summarised the key points in the first interview to the participants and double-checked whether I understood and interpreted them correctly, which enhanced the credibility of my interpretation.

In addition, I acknowledge the importance of reflecting the researcher's position in this study. The experiences and background of researchers influence the research in various stages, such as the literature review process, the choice of research design and the dynamics between researchers and participants, especially in the qualitative method.

First, most of the interviews in this study were undertaken in English, which was the language popularly used among international students. At the time the study was conducted, I was also an international student in the UK. The shared background and language position connected me with the participants. I was regarded as an 'insider' in this community of non-UK students, which helped me to gain some trust and build rapport with them. This helped the students to feel more comfortable in discussing some sensitive issues such as social grouping on campus.

Moreover, I was seen, to some extent, as an 'outsider' to the participants as I did not belong to these students' cohort (i.e., taught masters' students) and did not share the same cultural/ national background with the students. This position could enable me to maintain an objective view when examining the investigated phenomenon.

Findings and discussion

Firstly, much in line with previous studies (Brown, 2009a, 2009b; Fincher & Shaw, 2009; Glass, Gómez & Urzua, 2014; Kwon, Hernandez & Moga, 2019; Rienties & Nolan, 2014; Rose-Redwood & Rose-Redwood, 2013), the data suggested three primary forms of social grouping in UK HE: (1) international students self-segregated with people from their country of origin (i.e., co-nationals), (2) grouping between people from perceived 'different' geographical regions of origin (e.g., 'Asian' and 'Western' students), and (3) grouping between international students and 'home' students (i.e., UK students) (see Figure 1 below).

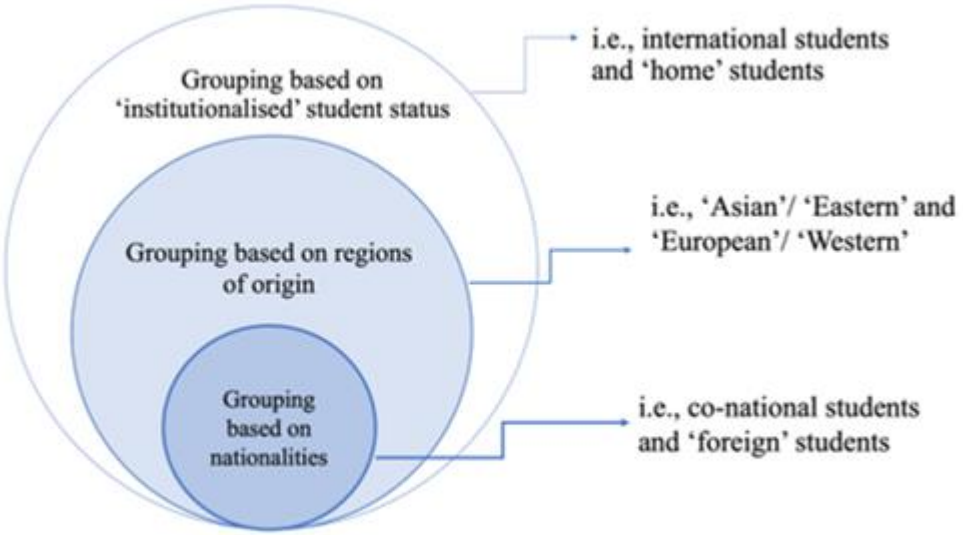


Figure 1. Three common forms of the social grouping of international students in UK HE

Secondly, region of origin is often regarded in a number of studies as the predictor of social grouping (Glass, Gómez & Urzua, 2014; Glass & Westmont, 2014; Rienties & Nolan, 2014; Wang, 2016). 'Asian' (or 'East Asian') students are usually represented in the intercultural communication literature as a 'disadvantaged' group who are expected to struggle in socialising with 'home' students, especially if the host country is 'Western' (e.g. the USA or the UK) (Glass, Gómez & Urzua, 2014; Glass & Westmont, 2014; Rienties et al., 2013; Slaten et al., 2016; Wang, 2016).

This study confirmed that students of the said group indeed struggled to socialise with students from the UK and other countries, but research findings presented another viewpoint. Figure 2 summarises contextual-level and individual-level factors that contributed to social grouping in British 'internationalised' HE.

In particular, the students with the lack of confidence in English communication skills were less likely to initiate conversations with students from other countries. The lack of confidence may impede communication with UK students and other international students and pushes them to retreat to their co-national groups, as found in previous research by Brown (2009a) and Hendrickson et al. (2011).

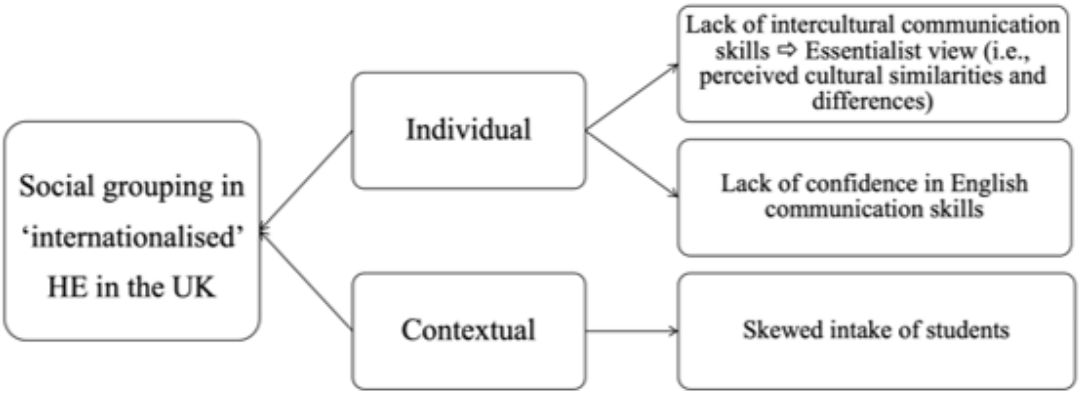


Figure 2. Summary of contributory factors of the social grouping of international students in UK HE

In particular, students who self-segregated with their compatriots shared that they often felt "not confident" (Interviewee 4, from China), "afraid… embarrassed" (Interviewee 3, from Hong Kong), "stupid… depressed" (Interviewee 5, from Kuwait), "shy" (Interviewee 6, from China) and "not convenient" (Interviewee 10, from China) when communicating in English with other international students.

"Not everyone can speak really well English so sometimes they might want to… try to hang out with foreigners, but at the same time they are afraid… they can't communicate well with them, they feel embarrassed when speaking English in front of a lot of different people…" (Interviewee 3, from Hong Kong, P1)

In addition, the anxiety intensified when they communicated with UK students. Some interviewees expressed that they felt "much pressured" (Interviewee 4, from China, P2) or even "scared" (Interviewee 5, from Kuwait, P2) when speaking English to UK students.

"…when they are native speakers, Welsh or English, maybe you will feel a bit of the pressure because they will give you words that are sometimes kind of difficult to understand..." (Interviewee 4, from China, P2)

At a contextual level, the skewed intake of students on the postgraduate programme (with a limited number of 'home' students) could make it innately challenging for international students to form strong relationships with their local colleagues.

More than half of the interviewees repeatedly explained that "…there aren't many English students in my class" (Interviewee 1, from Vietnam, P1) or "I rarely meet British students in the school" (Interviewee 13, from Japan, P1).

Small groups of British students staying together throughout the course

were frequently reported. As Interviewee 20, from Germany (P1) explained, most of the 'local' students seemed to retreat within their formerly built networks: "I think UK students, they came here and it's not new, like when you've been here before, you have yourself a group of friends."

Thirdly, this study found that the lack of intercultural communication skills was the main factors behind social grouping in 'internationalised' HE in the UK. International students, regardless of their regions of origin, generally had difficulties communicating with students perceived to be from the 'different' region, thus, tried to minimise or sometimes avoided cross-cultural interactions.

The students, to some extent, performed social essentialism, which is defined as the assumption that people can be 'classified into certain social categories (i.e. gender, race and nation)' (Rhodes, Leslie & Tworek, 2012, p. 13526). In this case, international students tended to distinguish other students based on 'institutionalised student status' (i.e., 'home' students and international students), regions of origin (i.e., 'East' and 'West') and nationalities. They attributed sets of characteristics as distinctive of members of a cultural group and homogenised that group and used labels such as "Asians", "Westerns", "locals" and "internationals" to refer to these student groups. For instance, 'Asian' students were perceived to be more timid, shy, and had relatively lower English proficiency whereas 'European' students were regarded to be more active, 'vocal' and were more fluent in English. Interactions with students from the 'other' cultural group were perceived to be stressful.

In particular, in the first few weeks, some interviewees tried conversating with students from the perceived 'culturally distant' groups, but the cultural difference in the communication style between 'the East' and 'the West' was regarded to be so challenging that it demotivated them from maintaining the relationships. Interactions with people from the 'other' group were expected to result in "awkwardness", as shown in the conversation with Interviewee 7 (from the USA, P1).

Interviewee 7: …it's so weird… today there was an Asian girl sat at our table and that's the most I can talk to, all year […] I feel like people can just sit where they feel comfortable. No one want to put themselves in an awkward situation so…

Interviewer: Why do you think you will have an awkward situation?

Interviewee 7: […] OK, if I talk to them, they don't know what I'm saying because of the language barriers so it's like I don't really want to, you know… Then, if I talk to them, like I don't know anything about China, so it's like why… what do I have in common then?

Based on the 'initial' cross-cultural experience, international students tended to generalise some attributes to be generic for people from the 'other' cultural group. Interviewee 9 (from China), for instance, shared that she sometimes could not participate in discussion when there were 'Western'

students in the group since they were more vocal and individualistic. Meanwhile, Interviewee 19 (from Germany) insisted that 'Asian' students tended to be shyer and more withdrawn from the discussion (see Figure 3).

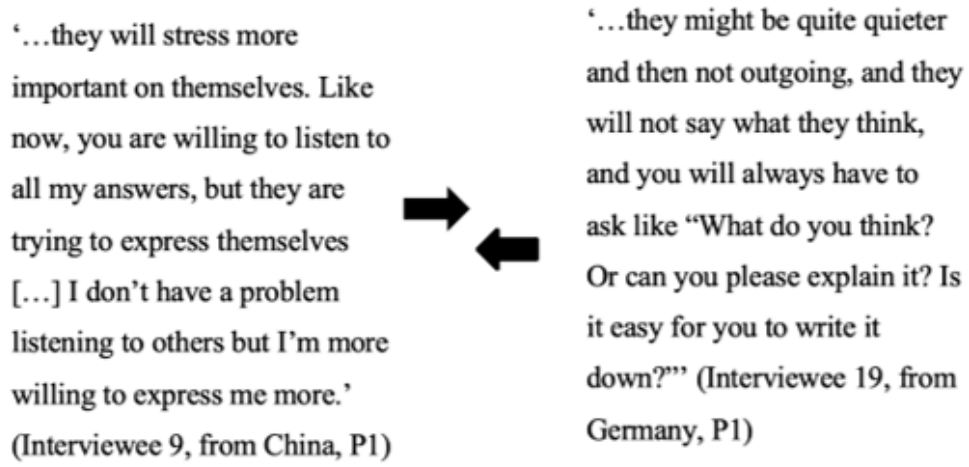


Figure 3. Perceived differences in communication styles of students from 'the East' and 'the West'

Although English was used as lingua franca, different ways of using and speaking English (i.e., speed and accent) were also seen as distinctive features of these two 'cultural' groups. For instance, 'European' students were assumed to have a higher level of English proficiency than 'Asian' students, due to the similarity between their 'native' languages.

"I think because maybe Asian people are like 'I know you from Asia, maybe you are also not good at English'. It can be like… because we speak the same level of English …" (Interviewee 4, from China, P1)

"I think the speed, how they speak and how they use English are different. […] Europeans speak quite fast, but I think it's because they share some structure or some words with their mother languages […] Like Germany, Spanish, English… they share some commonality, the language, and the speed is faster, the speed and the knowledge of the words…" (Interviewee 13, from Japan, P2)

In the first interview, Interviewee 22 used the word 'Asian' to describe her accent and created a sense of connection with me when she knew that I was also from a Southeast Asian country.

"…because my accent is like you… you know… Asian accent, you know, right? […] students from European countries, I think, they are more fluent than us, from Asian countries, even though English is not their first

language." (Interviewee 22, from Indonesia, P1)

Therefore, these students were more likely to feel pressured and lacked confidence when conversating with students from the 'West', while the latter appeared to regard conversations with 'Asian' students to be frustrating since they needed to accommodate their 'English' (see Figure 4).

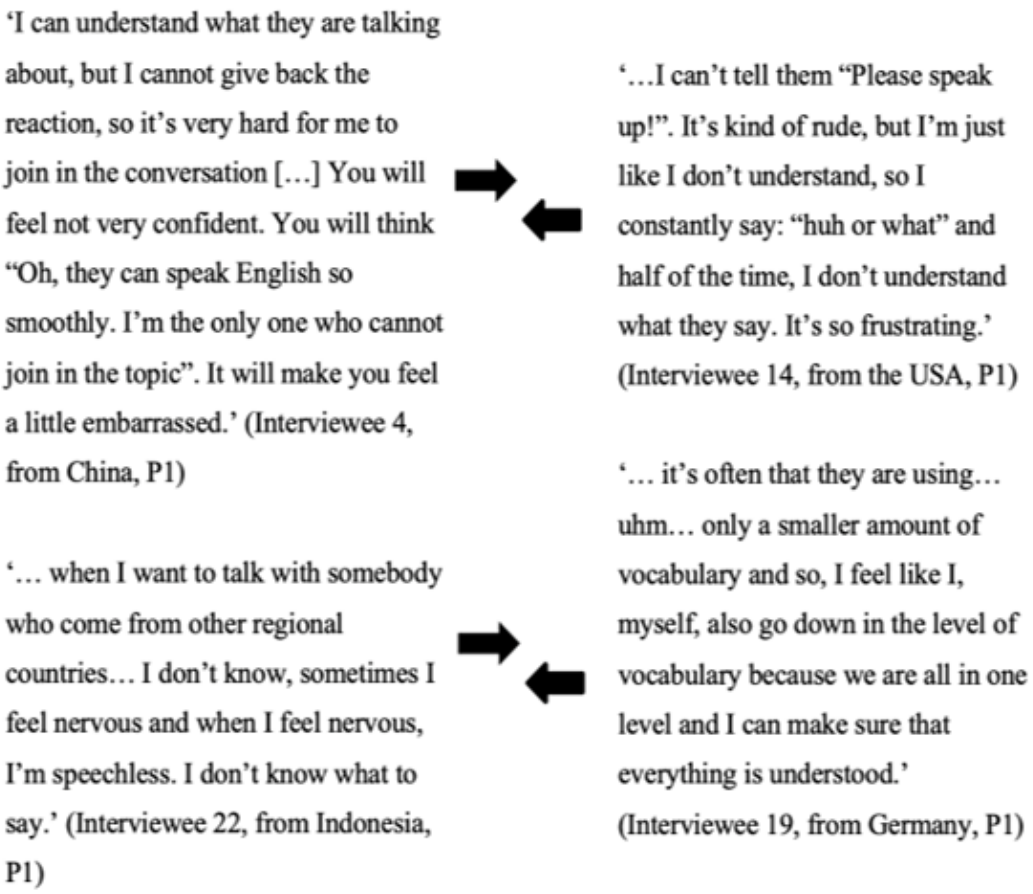


Figure 4. Perceptions of international students about challenges in cross-cultural communication of students from 'the East' and 'the West'

Interactions with students from the perceived 'different' cultural group were, therefore, either minimised or remained brief and superficial.

"I do approach them ['European' students], I do talk to them but like I said, it's just socialising, small talk, but we couldn't be more than that." (Interviewee 1, from Vietnam, P1)

Similarly, due to the comfort of linguistic familiarity, some students often self-segregated with their co-nationals.

"I kind of think it's important to have that, not to use it or act on it necessarily, but just to know if there is something going on, I can be on a space where I can speak my own language, I can express everything I want to express because English is kind of hard sometimes…" (Interviewee 21, from Germany, P1)

Phrases such as "same cultures, same background […] similar ideas" (Interviewee 9, from China), "similar thinking" (Interviewee 20, from Germany), "similar problem… same kind of language … same kind of perceptions or traditions" (Interviewee 21, from Germany) were usually stated as the main reasons. Cultural similarity connected the students with their compatriots, giving them more relevant topics to talk to each other, as Interviewee 15 (from China, P1) stated:

"I think because you are in an exchange experience, you may want someone close to you, someone from the same cultural background. Yes, from the same background. For me, I think I have more things to talk with them."

The construction of the cultural 'Self' and 'Other' could also be observed. On one hand, perceived cultural and linguistic similarity might 'pull' the students towards their perceived cultural groups, such as groups of co-nationals, 'Asian'/'Western' and 'international' students. On the other hand, cultural and linguistic differences 'pushed' the students further away from the students from 'other' cultural groups (e.g., 'foreign' students, 'Asian'/'Western', and 'home' students).

Each of these cultural groups had certain 'grouping' patterns, formed very early the sojourn (usually within the first few weeks) within the academic settings (i.e., specific 'seating' areas in classroom or lecture halls), which its group members were usually reluctant to break. Interviewees 13 and 14 even drew the seating 'maps' of their lecture hall and the seminar room (refer to Figures 5 and 6).

"Our classrooms are definitely divided in terms of nationality because generally in […] the right front or right middle, there are American, English and Germany, all together consistently. Then, you have in the left middle, you always have a few Americans and maybe a few Germans, and the rest are entirely like Chinese…" (Interviewee 14, from the USA, P1)

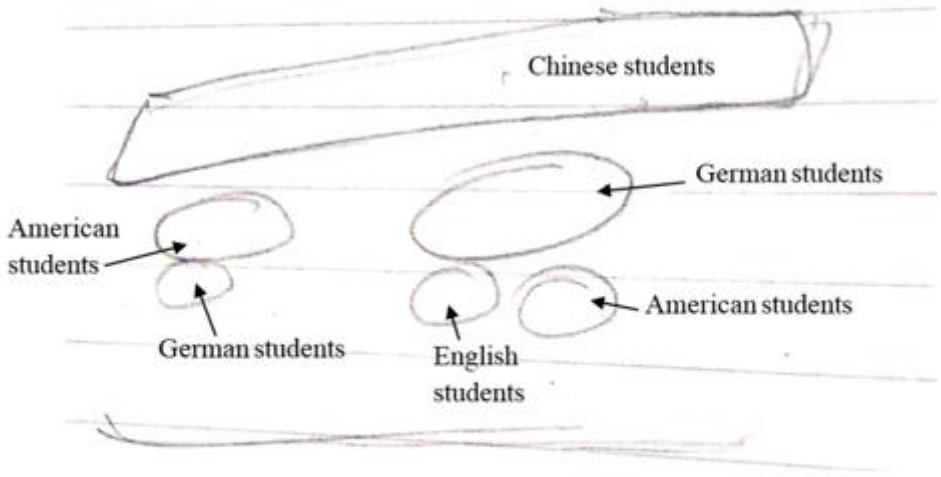


Figure 5. 'Seating map' of students in the classroom, drawn by Interviewee 14, from the USA, P1 (notes were added by the researcher).

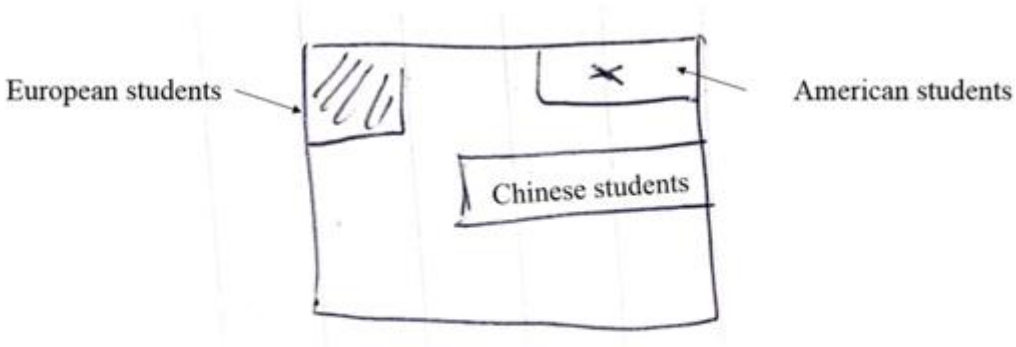


Figure 6. 'Seating map' of students in the lecture hall, drawn by Interviewee 13, from Japan, P2 (notes were added by the researcher).

"Can I write on this? This kind of group… like this is Chinese, and America, some Europeans there, that's the grouping, kind of." (Interviewee 13, from Japan, P2)

International students could sometimes be members of multiple groups, depending on the setting. For instance, Interviewee 1 (from Vietnam) reflected that the context influenced the grouping pattern. Within the academic contexts, such as classrooms, she often sat with 'Asian' students, but at her accommodation, she spent most time with her Vietnamese friends.

Interviewee 4, from the USA, shared that within the classroom setting, she usually sat with her co-nationals, but at her university accommodation, she befriended with her flatmates who were mostly international. They,

however, rarely interacted in class because they sat in different groups which were formed during the first few weeks.

"I mean I'm comfortable and familiar with X. We are flatmates, but still in our seminar class, she sits elsewhere, and I sit another place, we don't even interact there. It's just the patterns in the beginning…" (Interviewee 14, from the USA, P2)

Once set, these grouping patterns were maintained until the end of the academic year, with very few and minor modifications, or as in the words of Interviewee 14, "it's just hard to break the patterns once it's set in stone!". Many students also agreed on the importance of the first few weeks.

"I think that kind of group is already quite… grouped from the first day of this course. I think at the end of September, there was a welcoming meeting or something, if they get to know each other at that day, they continue to meet over the course of their study." (Interviewee 13, from Japan, P2)

"I think the first 2-3 months are quite important for you to settle down with people, to life here. You quite develop a kind of pattern, one kind of pattern, in the next few months." (Interviewee 16, from Taiwan, P2)

Some students, such Interviewee 14, from the USA, clarified in the second interview that she was less likely to change the grouping pattern since she worried this might offend their 'groupmates': "If I change my place to sit, they may find something strange, like 'Why not sit here?'".

In conclusion, findings showed that international students generally lacked intercultural communication skills to manage cross-cultural conflicts and adjust to different levels of English proficiency. If left unsolved, those conflicts might result in frustration in communication, which led to further avoidance of cross-cultural communication. The students, in general, need support and guidance to demystify the essentialist view of cultures and to create meaningful intercultural interactions.

Conclusion

This study was conducted to explore social grouping that international students might experience when study in a HE institution and the rationale behind it. The study found that there were generally three forms of grouping: 1.international students self-segregated with their compatriots; 2.grouping between people from perceived 'different' geographical regions of origin; and 3.grouping between international students and 'home' students (i.e., UK students).

These grouping patterns were formed early and remained unchanged until the end of the sojourn. The study found that factors for social groupings were 1.the skew intake of postgraduate students in British HE; 2. lack of confidence in English communication skills; and 3. lack of intercultural communication skills.

The study provided some theoretical contributions. In particular, many researchers in cross-cultural transition, taking the social psychological approach, often suggested regions of

origin as the main reason for social grouping and portrayed international students from East Asia and Southeast Asia as the ones struggling in socialising. However, this study offered a new perspective. Social grouping of international students in HE was complex and multi-layered (see Figure 1). Students could be members of multiple groups, depending on different contexts (e.g., classroom setting, 'international' accommodation and the university campus).

Students, regardless of their origins, could experience difficulties in cross-cultural communication. In particular, international students usually perceived that there existed two cultures: East (or Asia) and West. Cultural differences (e.g., different interests) between these two were perceived to be so significant that attempt to communicate with people from 'other' regional groups was anticipated to result in awkwardness and misunderstanding, even before any cross-cultural interactions happened. Therefore, although the students said they frequently mingled with their international peers, it was more likely that they grouped with friends perceived to be from the 'similar' regional group (e.g., 'Asia' or 'Europe') and minimised contact with those from the 'different' group.

As a result, this raises concern about the presentation of East Asian and Southeast Asian students in the cross-cultural transition study as the only so-called 'disadvantaged' group who usually struggle to maintain contact with internationals and host nationals (Rienties et al., 2013; Glass et al., 2014; Slaten et al., 2016). A new approach to study intercultural interactions between international students should be taken in which no student cohort should be treated as the 'only struggling one'.

Implications for practice

Research findings have implications for academic tutors, educators and HE. First, a series of workshops developing intercultural communication skills and improving the confidence in English communication skills should be provided early in the sojourn. Benefits of intercultural communication should be clearly conveyed to all students. For example, topics related to social integration and benefits of intercultural interactions can be disseminated through university media channels, such as university forums, magazines or newspapers and networking events.

Second, given the importance of the first few weeks of the academic year, pre-sojourn preparation tool kits can be offered to international students prior to the sojourn. For instance, online webinar, blogs and short videos of current international students and alumni sharing experiences of intercultural communication and social integration, some challenges and how they overcame them, could be posted.

Third, many UK HE institutions usually organise a separate orientation programme for international students. However, as grouping patterns were formed during this time, it is important that more opportunities for direct contact between international and 'home' students are created. Thus, the

programme can be restructured to include all students.

Fourth, 'buddy schemes', which focused on 'pairing' incoming students with 'home' students, could usefully be expanded to include international students, motivating mentoring among international students themselves.

Finally, it may be useful if HE institutions offer training to increase awareness of academic and teaching staff working with international students about some essentialist assumptions and how to mitigate those.

Limitations and suggestions for future research

First, the study was limited to a specific international student cohort in postgraduate taught programmes in a HE institution in the UK, aged around 20 to 30, with one-year overseas stay. These factors could possibly reduce the generalisability of the research findings. Future research could examine different student cohorts (e.g., students from different programmes, of different age groups and in other countries).

Second, the study was conducted on international students, thus, might not 'capture' the 'complete' picture of social grouping in UK HE. It may be useful if other researchers study viewpoints of different stakeholders, such as 'home' students and teaching staff.

Finally, qualitative research may be subject to credibility, such as the subjectivity of the researcher, and transferability, referring to the extent to which findings can be transferred to other contexts. Other researchers are, thus, recommended to employ a range of methods to further explore this complex phenomenon.

Acknowledgments

I would like to express my gratitude to all the students that contributed to the interview and informed the findings of this paper.

References

Atkinson, D. (1997). A critical approach to critical thinking in TESOL. *TESOL Quarterly, 31*(1), 71–94.

Biggs, J. B. (1996). Western misperceptions of the Confucian-heritage learning culture. In D. A. Watkins & B. J.B. (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 45–67). Hong Kong and Melbourne: Comparative Educational Research Centre and Australian Council for Educational Research.

Brown, L. (2009a). An ethnographic study of the friendship patterns of international students in England: An attempt to recreate home through conational interaction. *International Journal of Educational Research, 48*(3), 184-193.

Brown, L. (2009b). A failure of communication on the cross-cultural campus. *Journal of Studies in International Education, 13*(4), 439-454.

Brown, L. (2009c). The transformative power of the international sojourn: An ethnographic study of the international student experience. *Annals of Tourism Research, 36*(3), 502–521.

CAUT. (2016). CAUT Bulletin. Retrieved from http://www.caut.ca/bulletin/articles/2016/05/academics-warn-of-dangerous-dependency-on-internationalfees

Clarke, V. & Braun, V. (2014). Thematic analysis. In A. C. Michalos (Ed.), *Encyclopaedia of Quality of Life and Well-Being Research* (pp. 6626-6628). Netherlands: Springer.

De Vaus, D. A. (2001). *Research design in social research*. London: London : SAGE.

De Vita, G. (2002). Does assessed multicultural group work really pull UK students' average down? *Assessment & Evaluation in Higher Education, 27*(2), 153-161.

de Wit, H., Hunter, F., Egron-Polak, E. & Howard, L. (2015). *Internationalisation of Higher Education*. Paper presented at the European Parliament, Brussels. http://www.europarl.efuropa.eu/RegData/etudes/STUD/2015/540370/IPOL_STU(2015)540370_EN.pdf.

DFATD (2014). Canada's International Education Strategy: Harnessing Our Knowledge Advantage to Drive Innovation and Prosperity. Retrieved from http://international.gc.ca/global-markets-marches-mondiaux/assets/pdfs/overview-apercu-eng.pdf

Durkin, K. (2008). The middle way: East Asian master's students' perceptions of critical argumentation in UK universities. *Journal of Studies in International Education, 12*(1), 38-55.

Fincher, R. & Shaw, K. (2009). The unintended segregation of transnational students in central Melbourne. *Environment and Planning A, 41*(8), 1884-1902.

Fincher, R. & Shaw, K. (2011). Enacting separate social worlds: 'International' and 'local' students in public space in central Melbourne. *Geoforum, 42*(5), 539-549.

Floyd, C. B. (2011). Critical thinking in a second language. *Higher Education Research & Development, 30*(3), 289-302.

Forest, J. J. F. & Altbach, P. G. (2006). *International handbook of higher education* (Vol. 1): Springer.

Forsey, M., Broomhall, S. & Davis, J. (2012). Broadening the mind? Australian student reflections on the experience of overseas study. *Journal of Studies in International Education, 16*(2), 128-139.

Gareis, E. (2012). Intercultural friendship: Effects of home and host region. *Journal of International and Intercultural Communication, 5*(4), 309-328.

Gil, N. (2014). International students in the UK: who are they really? Retrieved from https://www.theguardian.com/education/2014/oct/13/-sp-international-students-in-the-uk-who-are-they

Glass, C. R., Gómez, E., & Urzua, A. (2014). Recreation, intercultural friendship, and international students' adaptation to college by region of

origin. *International Journal of Intercultural Relations, 42*, 104-117.

Glass, C. R. & Westmont, C. M. (2014). Comparative effects of belongingness on the academic success and cross-cultural interactions of domestic and international students. *International Journal of Intercultural Relations, 38*, 106-119.

Guba, E.G. and Lincoln, Y.S. (1989) *Fourth generation evaluation*. Sage.

Hall, E. T. (1976). *Beyond culture*. New York: Doubleday.

Hall, S., Held, D. & McGrew, A. G. (1992). Modernity and its Futures.

Hampden-Turner, C., & Trompenaars, F. (2000). *Building cross-cultural competence*. Chichester, UK: John Wiley.

Hanassab, S. (2006). Diversity, international students, and perceived discrimination: Implications for educators and counselors. *Journal of Studies in International Education, 10*(2), 157-172.

Harper, S. R. & Hurtado, S. (2007). Nine Themes in Campus Racial Climates and Implications for Institutional Transformation. *New Directions for Student Services, 2007*(120), 7–24.

Harper, S. R. & Quaye, S. J. (2007). Student Organizations as Venues for Black Identity Expression and Development among African American Male Student Leaders. *Journal of College Student Development, 48*(2), 127–144.

Harrison, N. & Peacock, N. (2009). Interactions in the International Classroom. *Internationalisation and the student voice: Higher education perspectives*, 125.

Harrison, N. & Peacock, N. (2010). Cultural distance, mindfulness and passive xenophobia: Using Integrated Threat Theory to explore home higher education students' perspectives on 'internationalisation at home'. *British Educational Research Journal, 36*(6), 877-902.

Hendrickson, B., Rosen, D. & Aune, R. K. (2011). An analysis of friendship networks, social connectedness, homesickness, and satisfaction levels of international students. *International Journal of Intercultural Relations, 35*(3), 281-295.

Kelly, P. (2000). Internationalizing the Curriculum: For Profit or Planet? In S. Inayatullah & J. Gidley (Eds.), *The University in Transformation: Global Perspectives on the Futures on the University* (pp. 162–172). Westport: Greenwood.

Khoo, S.M. (2011). Ethical Globalisation or Privileged Internationalisation? Exploring Global Citizenship and Internationalisation in Irish and Canadian Universities *Globalisation, Societies and Education, 9*(3–4), 337–353.

Knight, J. (2004). Internationalization Remodeled: Definition, Approaches, and Rationales. *Journal of Studies in International Education, 8* (1), 5–31.

Knight, J. (2011). Five Myths about Internationalization. *International Higher Education 62*(Winter), 14–15.

Kwon, S. A., Hernandez, X. & Moga, J. L. (2019). Racial segregation and the limits of international undergraduate student diversity. *Race Ethnicity and Education, 22*(1), 59-72.

Malhotra, N. K. (2010). *Marketing research : an applied orientation* (6th ed.. ed.). Upper Saddle River, N.J., London: Pearson Education.

Montgomery, C. & McDowell, L. (2009). Social networks and the international student experience: An international community of practice? *Journal of Studies in International Education, 13*(4), 455-466.

Nashrawan, T. & Cox, A. (2016). International students' networks: a case study in a UK university. *Studies in Higher Education, 41*(1), 182-198.

Nisbett, R. E., Peng, K., Choi, I. & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*(2), 291-310.

Paul, R. (1982). Teaching critical thinking in the “strong” sense: A focus on selfdeception, world views, and a dialectical mode of analysis. *Informal Logic Newsletter, 4*(2), 2-7.

Paul, R. (1993). *Critical thinking: What every person needs to survive in a rapidly changing world* (3rd ed.). Robert Park, CA: Sonoma State University, Center for Critical Thinking and Moral Critique.

Peacock, N. & Harrison, N. (2009). “It’s So Much Easier to Go with What’s Easy” “Mindfulness” and the Discourse Between Home and International Students in the United Kingdom. *Journal of Studies in International Education, 13*(4), 487-508.

Pho, H. & Schartner, A. (2019). Social contact patterns of international students and their impact on academic adaptation. *Journal of Multilingual and Multicultural Development*, 1-14.

Pitts, M. J. (2009). Identity and the role of expectations, stress, and talk in short-term student sojourner adjustment: An application of the integrative theory of communication and cross-cultural adaptation. *International Journal of Intercultural Relations, 33*(6), 450-462.

Ray, M. & Solem, M. (2009). Gauging disciplinary engagement with internationalization: A survey of geographers in the United States. *Journal of Geography in Higher Education, 33*(1), 103–121.

Rhodes, M., Leslie, S. & Tworek, C. M. (2012). *Cultural transmission of social essentialism.* Paper presented at the The National Academy of Sciences.

Rienties, B., Beausaert, S., Grohnert, T., Niemantsverdriet, S. & Kommers, P. (2012). Understanding academic performance of international students: the role of ethnicity, academic and social integration. *Higher Education, 63*(6), 685-700. Retrieved from http://www.jstor.org/stable/41477908

Rienties, B., Nanclares, N. H., Jindal-Snape, D. & Alcott, P. (2013). The Role of Cultural Background and Team Divisions in Developing Social Learning Relations in the Classroom. *Journal of Studies in International Education, 17*(4), 332-353.

Rienties, B. & Nolan, E. (2014). Understanding friendship and learning networks of international and host students using longitudinal Social

Network Analysis. *International Journal of Intercultural Relations, 41*, 165-180.

Rizvi, F. (2011). Theorizing Student Mobility in an Era of Globalization. *Teachers and Teaching, 17* (6), 693–701.

Rose-Redwood, C. R. & Rose-Redwood, R. S. (2013). Self-Segregation or Global Mixing?: Social Interactions and the International Student Experience. *Journal of College Student Development, 54*(4), 413-429. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1018108&site=ehost-live

Schartner, A. (2015). 'You cannot talk with all of the strangers in a pub': a longitudinal case study of international postgraduate students' social ties at a British university. *Higher Education, 69*(2), 225-241.

Schartner, A. & Cho, Y. (2017). Empty signifiers' and 'dreamy ideals': perceptions of the 'international university'among higher education students and staff at a British university. *Higher Education, 74*(3), 455-472.

Slaten, C. D., Elison, Z. M., Lee, J. Y., Yough, M. & Scalise, D. (2016). Belonging on campus: A qualitative inquiry of Asian international students. *The Counseling Psychologist, 44*(3), 383-410.

Soria, K. M. & Troisi, J. (2014). Internationalization at home alternatives to study abroad: Implications for students' development of global, international, and intercultural competencies. *Journal of Studies in International Education, 18*(3), 261-280.

Teichler, U. (2010). Internationalising Higher Education: Debates and Changes in Europe In D. Mattheou (Ed.), *Changing Educational Landscapes. Educational Policies, Schooling Systems and Higher Education – a Comparative Perspective* (pp. 263–283). Dordrecht: Springer.

Tickle, L. (2014). Is the UK still the destination of choice for international students? Retrieved from https://www.theguardian.com/education/2014/may/20/international-students-uk-issues

Turner, Y. & Robson, S. (2007). Competitive and cooperative impulses to internationalisation: Reflecting on the interplay between management intentions and the experiences of academics in a British university. *Education, Knowledge and Economy, 1*(1), 65–82.

UNESCO. (2019). Outbound internationally mobile students by host region. Retrieved from http://data.uis.unesco.org/Index.aspx?queryid=172

Volet, S. E. & Ang, G. (2012). Culturally mixed groups on international campuses: An opportunity for inter-cultural learning. *Higher Education Research & Development, 31*(1), 21-37.

Wang, M. (2016). The Impact of Cultural Values on Chinese Students in American Higher Education. *Qualitative Report, 21*(4).

Warwick, P. & Moogan, Y. J. (2013). A comparative study of perceptions of internationalisation strategies in UK universities. *Compare: A Journal of Comparative and International Education, 43*(1), 102–123.

Wright, C. & Schartner, A. (2013). 'I can't… I won't?'International students at the threshold of social interaction. *Journal of Research in International Education, 12*(2), 113-128.

Wu, W. & Hammond, M. (2011). Challenges of university adjustment in the UK: a study of East Asian Master's degree students. *Journal of Further and Higher Education, 35*(3), 423-438. doi:10.1080/0309877X.2011.569016

Young, T. J. & Schartner, A. (2014). The effects of cross-cultural communication education on international students' adjustment and adaptation. *Journal of Multilingual and Multicultural Development, 35*(6), 547-562. doi:10.1080/01434632.2014.884099

A literature review of empirical studies concerning the International English Language Testing System test

Yiyin Wang

ABSTRACT: With the number of empirical studies on the International English Language Testing System (IELTS) examination growing exponentially in recent years, a review of the literature on its speaking component is considered timely. This article reviews empirical studies concerning four key elements of the IELTS speaking test, which derive from O'Sullivan's (2011, p. 262) validation model: the test system, the test taker, performance, and the scoring system. Finally, this article concludes with recommendations for various stakeholders. For test developers, it is suggested that future studies should focus on reconceptualizing the speaking construct, addressing test topic issues, researching rating scale revision and rating behaviours, as well as investigating test-takers' physical/physiological characteristics. For test administrators, there is an urgent need to improve the quality of the feedback system and Enquiry-On-Results (EOR)[1] services. For test preparation course instructors, it is encouraged to stress the importance of pre-task planning, anxiety management, willingness to speak and self-regulation strategies. IST teachers should also help candidates understand that IELTS interview questions can be valuable opportunities to demonstrate their language proficiency.

Key words: *IELTS, the test system, the test-taker, performance, the scoring system*

Contact: Y.Wang254@newcastle.ac.uk

Introduction

The International English Language Testing System (IELTS) is an international language test designed to "measure the language proficiency of people who want to study or work where English is used as a language of communication" (IELTS, 2007, p. 2). In 2015 alone, more than 2.7 million tests were conducted across 140 countries (O'Sullivan, 2018, p. 1). Simultaneously, it is estimated that over 110 IELTS-funded empirical studies (Roothooft & Ruth Breeze, 2019, p. 3; Nakatsuhara et al., 2017b, p. 3; Seedhouse & Morales, 2017, p. 3) have been undertaken to validate this test. Given the increasing global impact of IELTS in recent years, it is necessary to present an up-to-date review of empirical studies on its speaking component. Indeed, some reviewers have attempted to evaluate the IELTS speaking test (IST) from practical and theoretical perspectives (e.g., Roshan, 2013; Karim & Haq, 2014; Quaid, 2018; Li, 2019), but their work seems to share two limitations. Firstly, these test reviewing papers aim to evaluate the strengths and weaknesses of IST in terms of its format (Roshan, 2013, p. 122; Quaid, 2018. p. 3), testing conditions (Karim & Haq, 2014, p. 155; Li, 2019, p. 10), rating reliability (Roshan, 2013, p. 121-122, Karim & Haq, 2014, p. 155; Quaid, 2018. p. 4; Li, 2019, p. 10) and

[1] EOR service: It is a paid service to remark the candidate's IELTS test if they are dissatisfied with their results.

practicality (Karim & Haq, 2014, p. 155; Quaid, 2018, p. 8; Li, 2019, p. 11). As a result, only a limited number of empirical studies have been discussed. Secondly, although these authors have mentioned theoretical concepts, such as construct validity, face validity, criterion-related validity (Karim & Haq, 2014, p. 154; Li, 2019, p. 4), they (except Quaid, 2018) have not explicitly justified why these validity concepts were chosen. To address these two issues, this article utilizes O'Sullivan's (2011) validation framework to review empirical investigations on the IST. By adapting the well-known social-cognitive framework for validating speaking tests (Weir, 2005, p. 46), O'Sullivan presents this new model and asserts that "it comprehensively defines each of its elements with sufficient detail as to make the model operationalizable" (O'Sullivan, 2011, p. 262). This model consists of four elements, which helps to structure the discussion of IST empirical studies: the test system, the test-taker, performance and the scoring system. Finally, new directions for further empirical investigation are provided for various stakeholders.

The test system

Construct

Since the construct defines what theory the test is built upon (Principles of Good Practice, 2016, p. 22), it has received continuous attention in the field of language assessment. Recent years have witnessed an increasing interest in exploring the construct of interactional competence when assessing speaking (e.g., Plough, 2018; Plough et al., 2018; Kim, 2018). Interactional competence, defined as "the ability to co-construct interaction in a purposeful and meaningful way, taking into account sociocultural and pragmatic dimensions of the speech situation and event" (Galaczi & Taylor, 2018, p. 227), was discussed in a special issue of the Language Testing Journal (see Plough (2018) for details), which contains several studies from various contexts. Perhaps the most relevant among these is from Roever and Kasper (2018), who highlighted the 'co-constructed' nature of the oral proficiency test and contended that two interactional features (prefaces and pre-sequences) can be observed and used to distinguish candidates across different levels. These findings indicate that one possible way of expanding IST construct is to integrate interactional competence features, such as "topic management, turn management, interactive listening, break-down repair and non-verbal or visual behaviours" (Galaczi & Taylor, 2018, p. 227). This proposal is also supported by an earlier study by Nakatsuhara et al. (2011, p .1), which reveals that the speaking construct in Part 3 of the IST differs from that in Part 2 because it is associated with listening-into-speaking abilities.

One attempt has already been made in the direction of redefining the IST construct. Seedhouse and Morales (2017) proposed the addition of an extra section (Part 4) to the IST, in which candidates could ask questions to examiners. The results show evidence that the examiner-led format is preferable to encourage more interaction, and that most examiners agree that this should be sequenced

after Part 2. Seedhouse and Morales (2017, p. 38) further argue that Part 4 is important as it provides candidates with an extra opportunity to demonstrate their knowledge and examiners can gain an extra perspective on candidates' use of language by asking questions, which helps with rating in general. This suggestion reflects the co-constructed nature of interactional competence and helps to address the issue of lacking interactivity in IST, one weakness that has been criticised by Li (2019, p. 8) and Quaid (2018, p. 6).

The need to expand the IST construct is made even more pertinent by the current increase in the rapid technological development of testing delivery modes. Following this trend, IELTS partners have funded a series of revolutionary studies to compare the live test under both face-to-face and video-conferencing modes (Nakatsuhara et al., 2016; Nakatsuhara et al., 2017b; Berry et al., 2018, Lee et al., 2021). The overall findings reveal that the video-conferencing mode can generally be considered as a paralleled alternative to the face-to-face mode, as they result in similar test scores. Exploring the role that online communication plays is a significant step forward, which aligns with the recent revision of the Common European Framework of Reference for Languages (Council of Europe, 2018) to include online interaction. More importantly, online testing through video-conferencing modes has been minimally affected by the COVID-19 pandemic and demonstrates even greater significance Post-COVID-19 as provision of the face-to-face live test is still heavily restricted.

Test task

Several empirical studies regarding IST tasks have placed their research focus on the pre-task planning time, as this has been argued by many scholars to influence candidates' speaking performance (e.g., Foster & Skehen, 1996, p. 299; Yuan & Ellis, 2003, p. 23; Ellis, 2005, p. 1). The extent to which planning time affects IST candidates' performance was first examined by Elder and Wigglesworth (2006, p. 14), who examined two groups of learners (intermediate and advanced) and found that varied preparation time (zero, one-minute and two-minute) did not significantly influence candidates' fluency, accuracy and complexity regardless of their proficiency levels. They attributed anxiety as a cause for candidates' poor use of the one-minute preparation time. However, this null finding contradicts the results reported by Weir et al. (2006, p. 16), who maintain that candidates of three proficiency levels achieved significantly lower mean scores under a 'no planning' condition than with the official one-minute preparation time. Despite these differences, one common factor is that pre-task planning strategies play a significant role in the IST and should become a focus of teaching in test-preparation programmes.

As another important area, task topic is mostly investigated by Conversation Analysis (CA) studies. The results indicate that IST topics are organized in a constrained manner and this

organization has been quite effective in fulfilling its institutional goal. Seedhouse and Harris (2011, p. 69) labelled the organization of topics as the 'topic-based Q-A adjacency pair', in which the examiner introduces a question first and the candidate is required to provide an answer and develop a topic coherently. Despite its constrained nature, this 'topic-based Q-A adjacency pair' structure has been quite successful in maintaining high reliability of examiners' input while eliciting sufficient speech examples from candidates. According to Seedhouse (2019, p. 1), this is due to the dual personality of topics that can be distinguished as topic-as-script and topic-in-action. The former is defined as 'the scripted statement of topic on the examiner's cards prior to the interaction', whereas the latter pertains to 'how the topic is developed or talked into being during the course of the interaction' (Seedhouse, 2019, p. 3). This means that topics in the IST ensure that examiners provide homogeneous input when asking questions and enable candidates to produce heterogeneous output that can be awarded different scores.

However, questions have been raised about the IST topics as some of them may be excessively demanding and culturally inappropriate. For instance, Seedhouse and Harris (2011, p. 104) found that some IST topics, particularly those in Part 3, can be too challenging as candidates may not be able to answer them due to a lack of specialist knowledge or experience. Cultural bias is another issue that emerged from one exploratory study in Bangladesh. In this study, Khan (2006, p. 61) reported that some "topics reflect western concepts and patterns of interaction and are not culturally appropriate for local candidates". In the same vein, other scholars (e.g. Roshan, 2013, p. 124; Inoue, 2021, p. 55) have also maintained that some inappropriate topics in IST can disadvantage candidates from certain cultural backgrounds and should be dealt with more seriously to guarantee fairness for all test-takers.

Another rarely discussed question is the limited number of topics in the printed test booklet. As an example, due to the large testing population in China and the candidates' eagerness to share testing experience on social media apps such as IELTS Bro, Weibo and Wechat, it is possible to produce a copy of IST scripted questions. Inoue et al (2021, p. 17) also reported that "in China, the new test booklet is available pretty much publicly, not quite accurately, after a week it has come out". This may explain the phenomenon described in Brown and Taylor (2006, p. 18) that in certain regions, candidates can be familiar with examiners' topics and often prepare rehearsed answers. More seriously, the availability of speaking test topics to Chinese candidates jeopardizes the fairness of IST as this can put other test-takers in an unfavourable position. Therefore, urgent attention to this matter is required.

The test takers' characteristics

Gulcher (2003, p. 118) highlighted the importance of identifying test-takers' characteristics in validation studies and his list includes several traits that

can potentially affect candidates' test performance, such as "age range, gender, geographical location, L1, cultural factors that may impact on task/topic choice, and educational system". A more comprehensive list was synthesised by O'Sullivan and Weir (2020, p. 26), who characterised these traits into three general categories: physical/ physiological, psychological and experiential characteristics.

In the context of the IST, the majority of empirical studies have concentrated on the psychological features of test-takers, but the results are quite mixed. It is found that anxiety is negatively associated with IST performance (Woodrow, 2006) while both willingness to speak (Riasati, 2018) and self-regulation abilities (Mahjoob, 2015) are positively related to speaking performance. Furthermore, research also shows that some psychological characteristics, such as learning style (Morovat, 2014) and personality (Souzandehfar et al., 2014), are not determiners for IST scores. These empirical findings have strong pedagogic implications as they provided IST instructors with valuable suggestions: employing scaffolding skills and relaxation techniques to relieve anxiety (Woodrow, 2006, p. 324), changing classroom atmosphere and seating arrangement to increase students' willingness to speak (Riasati, 2018, p. 14) and exploring rehearsal, goal setting, planning strategies to improve candidates' IST performance (Mahjoob, 2015, p. 188).

Studies of test-takers' experiential characteristics mostly investigate candidates' examination experience and these reports have revealed some level of dissatisfaction towards the IST. In Cheng and Curtis' (2010, p. 116) study, interviewees confess that there is a disparity between the sub-skill score (e.g., the speaking score) and their perceived English proficiency, which may pose a threat to the validity of the IST. Test-takers' complaints regarding the IST can also be found in Hamid and Hoang's (2018) large-scale study with data collected from 377 former candidates from 52 countries. Some test-takers commented negatively on the Enquiry of Results (EOR) service and feedback system for lacking clarity and transparency (Hamid and Hoang, 2018, p. 14). Similar points were made by Pearson (2019b), who investigated perceptions towards the IST from candidates who failed to reach their expected language requirements. Pearson (2019b, p. 13) concluded that test-takers who 'failed' have a low level of trust in the single marking of the IST and there is a pressing need for IELTS partners to improve the quality of the feedback system. Future studies may explore a variety of feedback options by comparing their effectiveness from the perspective of test-takers.

A finding of note in this review was that to date, there has been no empirical research on the test-takers' physical and physiological characteristics. This may be due to the relatively smaller number of IST test-takers with physical needs (e.g. wheelchair accessibility) compared to their counterparts who do not have these problems. Nevertheless, this does not mean that accommodating test-takers' unique needs is insignificant. On the contrary, both Weir (2005, p. 53) and Košak-

Babuder et al. (2019, p. 55) emphasized the importance of investigating accommodation's impact on test construct. With an increasing number of countries introducing legal requirements to provide disadvantaged test-takers with equal opportunities in language testing, there is a need to compare the differences between the standard and accommodated versions of the IST.

Performance

The investigation of IST candidates' speaking performance mainly focuses on the relationship between test-takers' discourse and given scores. In particular, some researchers (e.g., Seedhouse 2012; Seedhouse et al., 2014) have aimed to address which speaking features in test-takers' discourse can distinguish IST proficiency levels. There is no easy answer to this question but one overall finding is that a cluster of features contributes to a given score (Brown, 2006a, p. 1; Seedhouse et al., 2014, p. 23). This means the IST score is an overall impression, which is not driven by one dominant feature. Such understanding can benefit rater training by guiding new trainee examiners to view the candidates' performance holistically and take a range of features into concern while rating. Evidence was found in Seedhouse and Satar's (in press, p. 34) report in which examiners are inclined to make rating decisions in a cumulative way rather than attend to single instances or features in the candidates' discourse.

More specifically, quantitative research methods have been used in studies to examine some pre-defined features in fluency (Brown 2006a; Seedhouse et al., 2014), coherence (Iwashita, 2015), lexical resources (Read & Nation, 2006; Iwashita, 2015; Brown 2006a) and grammatical range/accuracy (Brown 2006a; Seedhouse et al., 2014; Roothooft & Breeze, 2019), with results indicating that most of the features increase in the expected direction as the score increases in test-takers' discourse. However, the investigation into salient pronunciation features that distinguish candidates across various levels lagged behind the other criteria reviewed above. This may be due to the unreadiness of the pronunciation scale, as it was only in 2008 that the original 4-point scale was expanded into a 9-point scale to be in line with the other three marking criteria. Issac et al. (2015, p. 4) investigated the pronunciation features that discriminate levels between Bands 5 to 8 and found that, although pronunciation features do contribute to the rating, no one distinguishes across these levels. Another issue identified by Issac et al. (2015, p. 34) is that the in-between IELTS pronunciation band descriptors, i.e., Bands 5 and 7, were found to show limited practical value. Since these two bands are not explicitly articulated with specific features, examiners must consult the adjacent band descriptors to arrive at a scoring decision. This inconvenience indicates that much work is required to specify what salient features define band 5 and 7 performance.

Unlike quantitative researchers who rely on the pre-defined features, some qualitative analysts have used

Conversation Analysis to identify a series of salient features in the test-takers' discourse at different proficiency levels. For example, Seedhouse and Harris (2011, p. 69) found that when developing topics, high and low achieving test-takers show differences in "length of turn, topic trouble, engagement with the topic, coherence, use of lexical items and projection of identity". Building upon these findings, Seedhouse (2012, p. 21) further argued that high achievers intend to answer the question in greater depth, develop the topic more coherently, experience less trouble and repair, choose to use a wider range of less common vocabulary construct themselves more often as confident high-achievers on an international stage and produce longer turns in Part 2. These results have strong pedagogic implications for test-takers who do not have much prior IST experience. Understanding how high achievers typically engage in the IST can immediately benefit advanced learners because they can turn each question into an opportunity to showcase their oral proficiency (Seedhouse & Nakatsuhara, 2018) rather than answer those questions in an arbitrary way. In contrast to high achievers, their counterparts may benefit from these findings in the long-term. This means even though low achievers may understand what they are supposed to do in the IST, the use of this knowledge is restrained by their limited linguistic competence. This understanding could help them to set future goals and clear a path to reach a higher score in the long run.

The scoring system

Examiners' rating behaviours

Since the reintroduction of the IST in 2001 (Taylor, 2001, p. 2), researchers have paid closer attention to examiners' perceptions, experiences and marking behaviours when using the new rating scale. One extensive study of this kind is a worldwide survey by Brown and Taylor (2006, p. 18), which showed that both the new format and the rating scale work well and most examiners commented positively on their experience of using the revised IST. However, some examiners still expressed concerns with the rating, such as "the lack of flexibility in wording prompts and modifying content of prompts, a lack of specificity in the wording of the scales, a need for greater discrimination in assessing pronunciation" (Brown & Taylor, 2006, p. 18). Brown's (2006a, p. 1) seminal study investigated examiners' rating processes in the revised IST suggested a similar result, as it was found that most raters adhered to the scale closely while rating. However, two problems with the rating scale still need to be addressed. The first issue is concerned with the overlapping of fluency/coherence with lexical or grammatical resources as key indicators (e.g., discourse markers and connectives), which can also be regarded as lexical units and a trait of sentence complexity. Secondly, echoing Brown and Taylor's (2006) worldwide survey, Brown (2006a, p. 19) reported that "there was a common perception that the (pronunciation) scale did not discriminate enough".

Following Brown's seminal study (2006a), much of the empirical work

pays particular attention to examiners' rating of pronunciation and rating scale problems involved. For instance, after the pronunciation revision project (Develle, 2008), Yates et al. (2011, p. 1) investigated the assessment of pronunciation using the new scale and found that examiners generally comment positively on the use of the revised pronunciation scale, but that the rating variation is substantial. Furthermore, echoing Issac and colleagues' (2015, p. 34) call for a more distinguishable pronunciation rating scale, Yates et al. also stressed the urgent need to train examiners to distinguish adjacent pronunciation bands more accurately as it was reported that raters/examiners tended to downgrade Band 7 candidates to a Band 6 score.

New rating mode

The rapid advancement in video recording technology has enabled researchers to explore the possibilities of introducing new rating modes. One example is a study by Nakatsuhara et al. (2017a, 2020), who compared the examiners' ratings and perceptions in live, audio and video rating modes. The results reveal that both live and video ratings are significantly higher than audio rating scores, which challenges the current Enquiry on Result (EOR) system that is based on the assessment of audio recordings. Therefore, Nakatsuhara et al. (2017a, p. 46; 2020, p. 21) assert the video rating to be a better option than audio rating if the double marking scheme is to be introduced. Another study by Seedhouse and Satar (in press) investigated the examiners' video rating process using the VEO (Video Enhanced Observation) app. This app allows examiners to add specific criteria, scores and comments in the video at specific moments when rating decisions are made. As a result, Seedhouse and Satar (in press, p. 46-47) have summarised a list of salient features that examiners orient to while assessing fluency/coherence, lexical and grammar in IST. The above-mentioned studies have made significant contributions to our understanding of video rating, but two areas still need to be explored further. Firstly, not much comparison has been made between different video rating modes since Nakatsuhara et al. (2017a, 2020) focuses on comparing audio and video rating, while Seedhouse and Satar (in press) only focus on one type of video rating. Understanding examiners' rating of video-recorded tests under different circumstances can provide important information for incorporating video ratings into the double marking scheme in the IST. Little is known about the similarities and differences between examiners' rating behaviours in different rating modes, as empirical research related to rating setting is rather scarce (O'Sullivan & Weir, 2020, p. 41). Furthermore, although Nakatsuhara et al. (2017a, p. 45; 2020, p. 19) as well as Seedhouse and Satar (in press, p. 42), direct our attention to the visual information in the ratings, nonverbal features are not systematically examined. This is certainly an area of future interest as Plough et al. (2018, p. 434) argue that it is high time nonverbal behaviours should be investigated and incorporated into interactional

competence when defining the speaking test construct.

Conclusions and recommendations

The empirical studies reviewed so far have provided solid evidence to support the validity of IST but additional research still needs to be conducted. Therefore, recommendations for future research are provided from three perspectives taking each party's special interest into concern.

For IELTS speaking test developers

Firstly, reconceptualising the IST should be the focus of future revision projects. One possible way forward is to build on the construct of interactional competence to investigate a series of features that are not included in the current rating scale, such as abilities to raise questions (Seedhouse & Harris, 2017), nonverbal elements (Nakatsuhara et al., 2017a, 2020), listen-into speaking abilities (Nakatsuhara et al., 2011) and online interaction features (Nakatsuhara et al., 2016; Nakatsuhara et al., 2017b; Berry et al., 2018; Lee et al., 2021).

Secondly, two serious issues regarding IST topics still need to be addressed. One is the limited number of topics in the printed test booklet. Inoue et al. (2021, p. 17) have suggested using tablet computers to digitalize IST topics. The benefits of such integration are apparent, as the electronic version is more portable, economical and eco-friendly. Furthermore, the electronic-based script has a much larger topic pool that can be updated periodically. This option is made even more possible as some ISTs have been delivered through video-conferencing software (Nakatsuhara et al., 2016; Nakatsuhara et al., 2017b; Berry et al., 2018, Lee et al., 2021) which can potentially feature this built-in-function. Certainly, further research comparing the impact of paper and electronic scripts on examiners' rating is needed before the massive application of digital test booklet. The other issue is concerned with topic appropriateness, as some questions are reported to be either too intellectually challenging (Seedhouse & Harris, 2011, p. 104) or culturally biased (Khan, 2006, p. 78). Therefore, IELTS partners should increase transparency regarding the item writing process and report how topics have been reviewed to avoid IST content inappropriateness.

Further research on the revision of the rating scale and examiners' rating behaviours under various rating modes is required. Revising and developing the pronunciation rating scale is especially critical as studies (e.g., Issac et al., 2015, p. 4; Yates et al., 2011, p. 1) have pointed out its weakness of having limited practical value. In particular, the in-between band score descriptors, such as Band 3, Band 5 and Band 7, should be further specified as examiners often have to consult the adjacent band descriptors to make a scoring decision. This specification can benefit examiner training on rating pronunciation and promote future studies into the salient pronunciation features of test-takers' speech samples. Understanding how examiners behave differently under

various rating conditions is another area worthy of exploration. For instance, future projects that compare raters' behaviours in various video rating modes can not only benefit rater training, but also inform the double marking scheme of the IST.

Lastly, future empirical studies can benefit substantially from expanding their research scope. One potential area is to examine test-takers' physical and physiological characteristics. Future research may compare the standard IST with its accommodated version, which is designed for physically disadvantaged candidates. This comparison can be used to assess the effectiveness of the accommodated version and inform its further revision to ensure test fairness.

For IELTS speaking test administrators

More endeavours should be made to improve the quality of the feedback system and EOR service in the IST. The former has been questioned for being too general without providing analytical insight into specific aspects of speaking that contribute to candidates' underperformance (Pearson, 2019a, p. 8). The system fails to offer personal suggestions to test-takers as the feedback reports are generated by combining pre-stored templates. Unfortunately, no empirical studies have looked at the IST feedback system's impact on candidates' test preparation, which deserves more attention from IST administrators (Pearson, 2019b, p. 17; Hamid and Hoang, 2018, p. 16). As for the EOR service, candidates' performance is reassessed based on audio recordings, which have been found to be lower than the live or video rating in some studies (e.g., Nakatsuhara et al., 2017a, p. 4; Conlan et al. 1994). Therefore, it has been recommended that video ratings be deployed as a parallel alternative for live ratings because it accesses a wider speaking construct, which includes nonverbal features that are absent in audio ratings (Nakatsuhara et al., 2017a, p. 46). Indeed, more research is needed to compare video ratings and live ratings to evaluate the impact of these modes on examiners' rating behaviours.

For IELTS speaking test instructors

This review also informs IST teachers about what content to cover in their courses. This may include pre-task planning procedures (Elder & Wigglesworth, 2006, p. 21), anxiety management skills (Woodrow, 2006, p. 324), methods to promote candidates' willingness to speak (Riasati, 2018, p. 14) and self-regulation strategies (Mahjoob, 2015, p. 188) as empirical results indicate these techniques can affect candidates' performance. Furthermore, it should be made clear to candidates that IST questions can be valuable opportunities to showcase their competence (Seedhouse and Nakatsuhara, 2018) rather than tedious interview questions designed to elicit their responses.

References

Berry, V., Nakatsuhara, F., Inoue, C. and Galaczi, E.D. (2018) 'Exploring the use of video-conferencing technology to deliver the IELTS Speaking Test:

Phase 3 technical trial', *IELTS Partners: British Council, Cambridge Assessment English and IDP: IELTS Australia,* pp. 1-58.

Brown, A. (2006a) 'An examination of the rating process in the revised IELTS Speaking Test', *IELTS Research Reports 6,* pp. 41-69.

Brown, A. (2006b) 'Candidate discourse in the revised IELTS Speaking Test', *IELTS Research Reports 6,* pp. 71-89.

Brown, A. and Taylor, L. (2006). 'A worldwide survey of examiners' views and experience of the revised IELTS Speaking test', *Cambridge ESOL Research Note*s, 26, pp. 14–18.

Cambridge English. (2016) *Principles of Good Pr*actice, Available at: https://www.cambridgeenglish.org/Images/22695-principles-of-good-practice.pdf (Accessed: 12 February 2021)

Cheng, L. and Curtis, A. (2010) *English Language Assessment and the Chinese Learner,* New York: Routledge.

Conlan, C.J., Bardsley, W.N. and Martinson, S.H. (1994) 'A study of intra-rater reliability of assessments of live versus audio-recorded interviews in the IELTS Speaking component', *unpublished study commissioned by the International Editing Committee of IELTS*.

Council of Europe. (2018) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment; Companion Volume with New Descriptors,* Available at: https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989 (Accessed: 12 February 2021).

DeVelle, S (2008) 'The revised IELTS pronunciation scale', *Research Notes,* vol 34, pp. 36-38.

Elder, C. and Wigglesworth, G. (2003) 'An investigation of the effectiveness and validity of planning time in Part 2 of the IELTS Speaking Test', *IELTS research reports,* 6, pp. 1-28.

Ellis, R. (2005) *Planning and Task Performance in a Second Language* (Vol. 11), John Benjamins Publishing.

Foster, P. and Skehan, P. (1996) 'The Influence of Planning and Task Type on Second Language Performance', *Studies in Second Language Acquisition,* 18 (3), pp. 299–323.

Fulcher, G. (2003) *Testing Second Language Speaking.* London: Pearson, Longman.

Galaczi, E. and Taylor, L. (2018) 'Interactional Competence: Conceptualisations, Operationalisations, and Outstanding Questions', *Language Assessment Quarterly,* 15 (3), pp. 219–236.

Hamid, M.O., and Hoang, N.T. (2018) 'Humanising Language Testing', *TESL-EJ,* 22(1), pp. 1-20.

Karim, S. and Haq, N. (2014) 'An Assessment of IELTS Speaking Test', *International Journal of Evaluation and Research in Education (IJERE),* 3 (3), pp. 152–157.

Kim, H. (2018) 'What constitutes professional communication in aviation: Is language proficiency enough for testing purposes?', *Language Testing,* 35(3), pp. 403-426.

Khan, R. (2006) 'The IELTS speaking test: Analysing culture bias', *Malaysian Journal of ELT Research,* 2(1), pp. 20.

Košak-Babuder, M., Kormos, J., Ratajczak, M., and Pižorn, K. (2019) 'The effect of read-aloud assistance on the text comprehension of dyslexic and non-dyslexic English language learners', *Language Testing,* 36(1), pp. 51-75.

Inoue, C., Khabbazbashi, N., Lam, D. and Nakatsuhara, F. (2021) 'Towards new avenues for the IELTS Speaking Test: Insights from examiners' voices', *IELTS Research Reports Online Series*, 2, pp. 1-70.

Isaacs, T., Trofimovich, P., Yu, G., and Chereau, B. M. (2015) 'Examining the linguistic aspects of speech that most efficiently discriminate between upper levels of the revised IELTS pronunciation scale', *IELTS Research Reports Online Series,* 48, pp. 1-48.

Iwashita, N. and Vasquez, C. (2015). 'An examination of discourse competence at different proficiency levels in IELTS Speaking Part 2', *IELTS Research Reports,* 5, pp. 1-44.

Lee, H., Patel, M., Lynch, J. and Galaczi, E. (2021) 'Development of the IELTS Video Call Speaking Test: Phase 4 operational research trial and overall summary of a four-phase test development cycle', *IELTS Research Reports Online Series*, 1, pp. 1-48.

Li, J. (2019) 'An Evaluation of IELTS Speaking Test', *Open Access Library Journal,* 6(12), pp. 1-17.

Morovat, E. (2014) 'Effects of Reflectivity/Impulsivity on IELTS Candidates' Band Scores in the Speaking Module of the Test', *Procedia - Social and Behavioral Sciences,* pp. 1232–1239.

Nakatsuhara, F., Inoue, C., Berry, V. and Galaczi, E.D. (2016) 'Exploring performance across two delivery modes for the same L2 speaking test: face-to-face and video-conferencing delivery: a preliminary comparison of test-taker and examiner behaviour', *The IELTS Partners: British Council, Cambridge English Language Assessment and IDP: IELTS Australia,* pp. 1-67.

Nakatsuhara, F., Inoue, C. and Taylor, L. (2017a) 'An investigation into double-marking methods: comparing live, audio and video rating of performance on the IELTS speaking test', *The IELTS Partners: British Council, IDP: IELTS Australia and Cambridge English Language Assessment,* pp. 1-49.

Nakatsuhara, F., Inoue, C., Berry, V. and Galaczi, E.D. (2017b) 'Exploring performance across two delivery modes for the IELTS Speaking Test: face-to-face and video-conferencing delivery (Phase 2)', *The IELTS Partners: British Council, Cambridge English Language Assessment and IDP: IELTS Australia,* pp. 1-74.

Nakatsuhara, F., Inoue, C., and Taylor, L. (2020) 'Comparing Rating Modes: Analysing Live, Audio, and Video Ratings of IELTS Speaking Test Performances', *Language Assessment Quarterly,* pp. 1-24.

Read, J. and Nation, P. (2006) 'An investigation of the lexical dimension of the IELTS speaking test', *IELTS Research Reports,* 6, pp.207-231.

O'Sullivan (2011) 'Language testing', in Simpson, J. (ed.) *The Routledge Handbook of Applied Linguistics*: Taylor and Francis, pp. 259-273.

O'Sullivan, B. (2018) 'IELTS (International English Language Testing System)', in John I. Liontas, Tesol International Association, and Margo DelliCarpini (ed.) *The TESOL Encyclopedia of English Language*

Teaching. Hoboken, NJ, USA: John Wiley and Sons, Inc. pp. 1–8.

Weir, C., O'Sullivan, B. and Horai, T. (2006) 'Exploring difficulty in speaking tasks: An intra-task perspective', *International English Language Testing System (IELTS) Research Reports,* 6, pp. 1-42.

Pearson, W. S. (2019a) 'Critical perspectives on the IELTS test', *ELT Journa*l, 73(2), pp. 197-206.

Pearson, W.S. (2019b) ''Remark or retake'? A study of candidate performance in IELTS and perceptions towards test failure', *Language Testing in Asia,* 9 (1), pp. 1-20.

Plough, I. (2018) 'Revisiting the speaking construct: The question of interactional competence', *Language Testing,* 35 (3), pp. 325–329.

Plough, I., Banerjee, J. and Iwashita, N. (2018) 'Interactional competence: Genie out of the bottle', *Language Testing,* 35 (3), pp. 427–445.

Quaid, E.D. (2018) 'Reviewing the IELTS speaking test in East Asia: theoretical and practice-based insights', *Language Testing in Asia,* 8 (1), pp. 1-9.

Rabab'ah, G. (2016) 'The Effect of Communication Strategy Training on the Development of EFL Learners' Strategic Competence and Oral Communicative Ability', *Journal of Psycholinguistic Research,* 45 (3), pp. 625–651.

Riasati, M. J. and Xian-han Huang, Y. (2018) 'Willingness to speak English among foreign language learners: A causal model', *Cogent Education,* 5 (1), pp. 1-17.

Roever, C. and Kasper, G. (2018) 'Speaking in turns and sequences: Interactional competence as a target construct in testing speaking', *Language Testing,* 35 (3), pp. 331–355.

Roothooft, H. and Breeze, R. (2019) 'Investigating the development of grammatical range and accuracy at different proficiency levels in the IELTS Speaking test', *IELTS Research Reports Online Series,* 36, pp. 1-36.

Roshan, S. (2013) 'A Critical Review of the Revised IELTS Speaking Test', *International Journal of English Language Education,* 2 (1), pp.120-127.

Seedhouse, P. (2012) 'What kind of interaction receives high and low ratings in Oral Proficiency Interviews?', *English Profile Journal,* 3(1), pp. 1-24.

Seedhouse, P. (2019) 'The dual personality of 'topic' in the IELTS Speaking Test', *ELT Journal,* 73(3), pp. 247-256.

Seedhouse, P. and Harris, A. (2011) 'Topic development in the IELTS Speaking Test', *IELTS Research Reports,*12, pp. 69-124.

Seedhouse, P., Harris, A., Naeb, R., and Üstünel, E. (2014) 'Relationship between speaking features and band descriptors: A mixed methods study', *The IELTS Research Reports Online Series,* 30, pp. 1-30.

Seedhouse, P. and Morales, S. (2017) 'Candidates questioning examiners in the IELTS Speaking Test: An intervention study', *IELTS Research Reports Online Series,* 5, pp. 1-43.

Seedhouse, P. and Nakatsuhara, F. (2018) *The discourse of the IELTS speaking test: Interactional design and practice*. Cambridge University Press.

Seedhouse, P. and Satar, M. (forthcoming) 'VEO IELTS Project

Report', *The IELTS Research Reports Online Series,*

Souzandehfar, M., Soozandehfar, S.M.A., Farsi, M. and Sharif, M. (2014) 'Which Personality Trait Performs better on IELTS Speaking Test? Extroverted or Introverted?', *Advances in Environmental Biology,* 8(6), pp.2159-2169.

The International English Language Testing System. (2007) *IELTS Handbook 2007*, Available at: https://www.ielts.org/pdf/IELTS_Handbook.pdf (Accessed: 12 February 2021)

Taylor, L. (2001) 'Revising the IELTS speaking test: Developments in test format and task design', *Cambridge ESOL Research Note*s, 5, pp. 2-5.

Weir, C.J. and O'Sullivan, B. (2020) 'Research issues in testing spoken language', in Taylor, L.B. and Saville, N. (ed.) *Lessons and legacy: a tribute to professor Cyril J Weir* (1950-2018), [online] Available at: https://www.cambridgeenglish.org/Images/582822-silt-volume-50.pdf, pp. 23-53.

Weir, C.J. (2005) *Language Testing and Validation*. London: Palgrave Macmillan UK.

Woodrow, L.J. (2006) 'A Model of Adaptive Language Learning', *The Modern Language Journal,* 90 (3), pp. 297–319.

Yates, L., Zielinski, B. and Pryor, E. (2011) 'The assessment of pronunciation and the new IELTS pronunciation scale', *IELTS Research Reports*, 12, pp. 1-46.

Yuan, F. and Ellis, R. (2003) 'The Effects of Pre-Task Planning and On-Line Planning on Fluency, Complexity and Accuracy in L2 Monologic Oral Production', *Applied Linguist*ics, 24(1), pp. 1-27.