CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

Research Matters

Issue 34 / Autumn 2022

Proud to be part of the University of Cambridge

Cambridge University Press & Assessment unlocks the potential of millions of people worldwide. Our qualifications, assessments, academic publications and original research spread knowledge, spark enquiry and aid understanding.

Citation

Articles in this publication should be cited using the following example for article 1: Carroll, M., & Constantinou, F. (2022). Learning loss in the Covid-19 pandemic: teachers' views on the nature and extent of loss. *Research Matters: 34,* 6–26.

Credits

Reviewers: Tom Benton, Lucy Chambers, Daniel Morrish, Joanna Williamson and Carmen Vidal Rodeiro

Editorial and production management: Lisa Bowett

Additional proofreading: Alison French

Cambridge University Press & Assessment is committed to making our documents accessible in accordance with the WCAG 2.1 Standard. We're always looking to improve the accessibility of our documents. If you find any problems or you think we're not meeting accessibility requirements, please contact our team: Research Division, ResearchDivision@cambridge.org

If you need this document in a different format contact us, telling us your name, email address and requirements and we will respond within 15 working days.

All details are correct at the time of publication in September 2022.

Contents

Foreword

“Lest we forget.” As a nation, England in 2020 ramped up its research on young people and the formal data requirements of schools – ranging from interviews and surveys of pupils and teachers to national submission of data on pupil absences. Researchers turned to the situations of young people and the issues being experienced by schools. Surveys were started, stats on school attendance were collected and, as young people returned to school, assessments were forensically examined for the patterns of loss and impact – we rightly sought understanding of the disruption to education, the impact on wellbeing and the pressured realities of schools. The findings of all studies in England converge on a single view of the disruption – while a few children benefited from the processes of remote learning and time in the home, the pattern of impact for the vast majority of the children in the country is negative, highly individualised and variable. This is the worst form of problem to respond to if you are a policy maker. It is fantastically hard to devise a means of supporting young people when the pattern of impact is so distributed and varied. The regional patterns in attainment emerging from the resumption of public exams indicates some systematic impact associated with deprivation, which we hope will enable government and agencies to consider how best to target support. The worry is that, as we try to resume “normality”, it would be all too easy to forget the experience of young people who were obliged not to attend school during the pandemic. Younger children entering primary school are likely to have experienced restricted social interactions and lack of participation in structured Early Years provision. Prior research tells us that we should not lapse into an “everything back to normal” sentiment and thus underestimate the long-term impact of any of these issues. But at least very young children have most of their compulsory education ahead of them, with more time to address issues of cognitive and social and emotional development. By contrast, those whose upper secondary education was adversely affected may have been unsuccessful in their efforts to progress to higher education or other destinations but now have no entitlement to continued fully funded education and training. This is the territory of “missing figures”, “forgotten third” and young people alienated from education and with little “voice”. It cannot be a “return to normal” in our support for any of these affected groups; they require us not to forget them. We need to research them for the purpose of ***action*** – to understand their circumstances and individual and collective experiences, and to put in place effective evidence-based support. And quickly.

Tim Oates, **CBE** Group Director, Assessment Research and Development

Editorial

The first article in this issue, by Matthew Carroll and Filio Constantinou, is another contribution to the large amount of research on the impact of the COVID-19 pandemic on education, as perceived by teachers. It is particularly noteworthy in its global coverage, including teachers from 38 countries. The findings about the perceived amount of "learning loss" are similar to those obtained from more quantitative studies based on assessment results, but teachers' comments about what was "lost" provide an opportunity to explore which skills were most affected and highlight the importance of variability in the extent of loss.

Our second article, by Joanna Williamson and Tom Benton, is more technical but right at the heart of assessment: how to maintain or link standards from one version of a test or exam to another. Making use of pairs of assessments that differed only in the "cover sheet" but were otherwise identical, they evaluated different statistical methods of linking the mark scales in the unusual condition where we happen to know the "right answer". The results were quite sobering but not unexpected – no single method was consistently better, and all methods could offer useful information but also could lead to incorrect conclusions about relative difficulty.

Our third article, by Chris Jellis, is an interesting exploration of a large data set from Cambridge CEM's BASE assessment. It shows what we can learn from assessment of very young learners just starting primary school – both what they can do when they arrive at school and what progress they make in their first year in a variety of key areas such as word and number recognition, and vocabulary.

Our fourth article, by Martin Johnson, reflects on the concept of "recovery curricula" developed in response to educational disruption. This is an area that Cambridge had been involved with prior to the pandemic, but obviously has now become particularly salient. The article considers how recovery curricula have been defined in the research literature and notes the lack of evidence (so far) for the effectiveness of any particular examples of where a recovery curriculum has been implemented.

The final article, by Joanna Williamson and me, is a bit of a departure from our usual fare. We investigated whether there are any systematic differences in the exam results of groups of students with different categories of surname and found a small effect in line with our hypothesis: average grades of candidates with "occupational" surnames were slightly lower than those in other categories. The article notes some possible explanations that have been proposed in the research literature (for other surname-related differences) but concludes that these are highly complex matters where findings should be interpreted with caution.

Tom Bramley Director, Research Division

Learning loss in the Covid-19 pandemic: teachers' views on the nature and extent of loss

Matthew Carroll and Filio Constantinou (Research Division)

Introduction

In 2020, schools around the world were closed in response to the Covid-19 pandemic. Initially, school closures were considered to be emergency measures to control the spread of the virus, but as the pandemic progressed it became evident that the disruption would be longer lasting. Where possible, teaching moved online to ensure education could continue in some form, prompting rapid changes to teaching and learning. Closures persisted for many weeks, if not months, with closures still in place in some parts of the world in early 2022, almost two years after the start of the pandemic. In some cases, schools reopened but were forced to close again in response to increased infection rates. Once schools were able to reopen, face-to-face teaching could be re-established (for some, if not all, students), but high absence rates and Covid control measures caused continued disruption to 'normal' schooling. Although at the time of writing most schools have reopened, the impacts of the disruption are ongoing and will be felt for some time still to come.

When considering the effects of the disruption to education, a major focus of attention has been "learning loss". In the context of Covid, learning loss is typically understood to be the "gap" between post-pandemic attainment (as observed by teachers or measured by tests) and that which would be expected *had it not been for the pandemic* (e.g., Newton, 2021; Renaissance Learning & the Education Policy Institute, 2021). To that end, it could represent either absolute loss (i.e., students have forgotten things they had previously learned) or relative loss (i.e., less progress has been made than in a typical year). Various attempts have been made to understand and quantify this loss (e.g., Donnelly & Patrinos, 2021; Engzell, Frey & Verhagen, 2021; König & Frey, 2022; Newton, 2021). To measure learning loss, one approach is for students to take progress tests (usually in mathematics and the student's first language) that have been standardised to a pre-pandemic population, such that any discrepancies from expected scores can be assumed to relate to Covid disruption (e.g., Renaissance Learning & the Education Policy Institute, 2021; Rose et al., 2021). Results from these studies have been reasonably consistent, with most estimates in the range 1–2 months "lost" (see figures collated by Newton, 2021), although with some studies indicating greater losses (e.g., Dorn et al., 2021). Other patterns identified from this approach include greater losses for disadvantaged students, regional variation in losses, and greater impacts on younger children (e.g., Renaissance Learning & the Education Policy Institute, 2021; Twist, Jones & Treleaven, 2022).

Valuable insights into learning loss have been gained from studies using standardised testing, but there are shortcomings to this method. Notably, standardised tests have been developed to measure specific learning areas, so results can only tell us about those areas. Further, the sample of students taking the test may be relatively small, and potentially unrepresentative of the wider population, thus making interpretation of the results challenging. Accordingly, an alternative approach is to survey or interview teachers and other education professionals, to gather expert opinion on what, and how much, has been lost (e.g., Chen et al., 2021; Sharp et al., 2020). Although this is inherently subjective and is also likely to rely on small samples, it can provide a more nuanced view of what may have been lost, and still permits a degree of quantification. That is not to say that such qualitative approaches are better than those based on standardised tests, but by allowing us to look beyond the amount of loss in a restricted range of topics, they can help us to better understand the nature of loss. Indeed, by considering both the amount *and* nature of loss, we should be better placed to understand the impacts on learning and, hopefully, better placed to help students recover what was lost.

In this study, we sought to understand more about learning loss by taking the latter of the approaches described above: we carried out a survey of teachers to gather opinions on the impacts of the pandemic on education. In doing this, we had several key aims. First, we wished to gather views from a diverse range of teaching settings, to uncover the breadth of impacts. Second, we aimed to make no assumptions about the nature or magnitude of any impacts; if respondents felt their students were ahead in some areas but behind in others, or even if they felt there were no impacts, opportunities were provided to report such observations. Finally, we hoped to gather insights that could inform practice.

Methods

Survey design and sample selection

Given the widespread nature of the Covid-19 pandemic, affecting the whole world and all stages of the education system, we wanted the survey to reach teachers from a diverse range of settings: focusing on a single country, one age group, or one school type might miss important aspects of the story. To achieve this, we collaborated with Cambridge CEM. CEM provides baseline and entrance tests to schools around the world, working with both state and independent sectors, and offering tests from early years up to upper secondary level. Hence, by surveying teachers from schools that use CEM tests, we could achieve the diversity of response desired.

We developed a survey to cover four main areas, all focused on teachers' experiences of teaching during the pandemic. The first area was impacts on students; this is the source of the results described in this article. The other major areas were impacts on teachers, experiences of remote teaching, and adaptations to teaching methods. Results from these other areas will be made available at a later date.

The survey primarily consisted of short, closed response questions, such as Likert scales or tick boxes. In most cases, optional free text boxes were provided beneath the main question to allow participants to provide further information. This approach was used to maximise the amount of data generated, by making it simple for participants to answer; any aspects that might take more time were entirely optional. Hence, the main role of the survey was to generate quantitative response data, but with the potential to also generate qualitative data.

Questions were developed over several drafting cycles. Once a final draft was created, questions were entered on to an online survey platform. This draft survey was piloted by two research colleagues with teaching backgrounds and by one current teacher. The pilot aimed to identify any areas where questions were unclear or which would be difficult to answer. Changes were made in response to pilot feedback, leading to the final version being created. This final version was put through Cambridge University Press & Assessment's research ethics approval process and reviewed by the data protection team to ensure all relevant ethical and legal standards were met.

Following the development and approval process, invitations were sent to the named contacts of all schools that use CEM tests or receive CEM marketing. Along with being invited to take the survey themselves, recipients were told they could pass the invitation to colleagues in their school if they wished. This sampling process was designed to generate as large a response as possible to take advantage of the breadth and diversity of schools that work with CEM. We acknowledge, however, that certain school types could end up over- or under-represented; the final sample composition is presented in the following Results section. Invitations to take part were sent on 23 April 2021, and the survey was open to responses for two months.

Data processing and analysis

Once the survey was closed, data was downloaded for analysis offline. Contact details were removed and school names were converted to pseudonyms, so that no individual or school could be identified during analysis. Respondents who had not consented to take part and those who had only answered the earliest contextual questions were removed, leaving data from 404 anonymous respondents.

Before analysing the data, several grouping variables were constructed to allow us to make comparisons of interest. First, to examine geographical variation in responses, respondents were split into either "UK" or "rest of the world" (hereafter, "RoW"). Approximately half of respondents were from the UK, but no other individual country had enough respondents to permit conclusions to be drawn. Hence, grouping respondents from all other countries permitted some exploration of geographical variation, focusing on how the UK differed from other locations. Next, respondents were split based on school type, with comparisons made between independent and state schools. Only UK schools were considered for this grouping, as almost all international schools were independent. The final grouping was based on the age of pupils; schools were classed as either primary (teach ages up to 11-12), secondary (teach ages from 11-12 and above), or mixed (teach a wider range of age groups). Most were either primary or secondary, so comparisons were made between these two groups. These groupings therefore permitted comparisons that could highlight differences between geographical regions, school types and age groups.

Data analysis focused on descriptive summaries. For closed items, we calculated simple counts and percentages of each response; we did this across all respondents, and separately for the groupings described above. We also read and summarised all free text responses, identifying broad themes discussed. Note that as free text responses were not mandatory, this analysis was carried out to provide context to support interpretation of the closed questions, rather than as a full, formal content analysis.

Results

Sample composition

There were 404 respondents, of whom 199 (49.3 per cent) were from the UK and 205 (50.7 per cent) were from other countries. Outside of the UK, the largest groups of respondents were from China (30; 7.4 per cent), India (30; 7.4 per cent), Italy (13; 3.2 per cent), Malaysia (13; 3.2 per cent), Switzerland (12; 3.0 per cent), UAE (11; 2.7 per cent) and Qatar (10; 2.5 per cent). In total, 38 countries were represented. Respondents came from 198 schools but were unevenly distributed among them; the largest number of respondents from a single school was 23, while 149 schools had only a single respondent. 79.5 per cent of respondents overall said their school did not receive state funding, but this was much greater in RoW than in the UK (92.7 per cent RoW, 65.8 per cent UK). Most respondents were from secondary schools (77.4 per cent overall), with 14.8 per cent from primary schools and 7.8 per cent from schools that fell into neither main category. Hence, the survey achieved the broad diversity of respondents hoped for, but we acknowledge that the sample is skewed towards certain conditions (i.e., UK schools, independent schools and secondary schools).

Considering the respondents themselves, almost all were teachers, with nearly every respondent saying that they were a classroom teacher or someone with oversight of teaching (e.g., school principals); over 96 per cent of respondents had clear teaching roles, with the remaining respondents having roles such as examinations officer or Special Educational Needs (SEN) co-ordinator. A larger-than-expected proportion were highly experienced, with 37.6 per cent having taught 21 years or more, 20.0 per cent having taught 16-20 years, 16.6 per cent having taught 11-15 years, 14.9 per cent having taught 6-10 years, and only 10.9 per cent having taught 5 years or fewer. This was also reflected in seniority of respondents, with 32.2 per cent having a senior leadership role and 34.7 per cent having another leadership role, with only 28.7 per cent of respondents not having a leadership role. Finally, concerning the subjects taught, 36.4 per cent of respondents said they taught humanities, 30.7 per cent taught sciences, 28.0 per cent taught English, 25.5 per cent taught mathematics, and 12.9 per cent taught creative subjects (note that respondents could select multiple subjects here). Hence, the sample of respondents showed diversity in the level of experience and in the subjects taught.

Estimates of learning loss

Respondents were asked “How far ahead or behind in their curriculum learning do you feel most of your students are at the moment, compared to in a ‘typical’ year?” Responses are plotted in Figures 1 and 2; the counts and percentages underlying the figures are presented in Appendix Table 1.

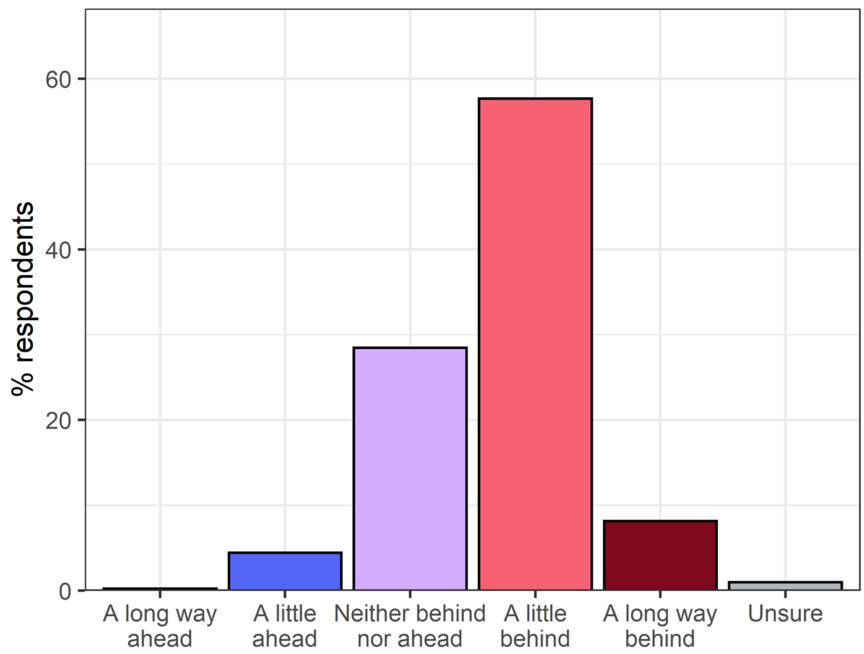

Figure 1: Overall responses to the question “How far ahead or behind in their curriculum learning do you feel most of your students are at the moment, compared to in a ‘typical’ year?”

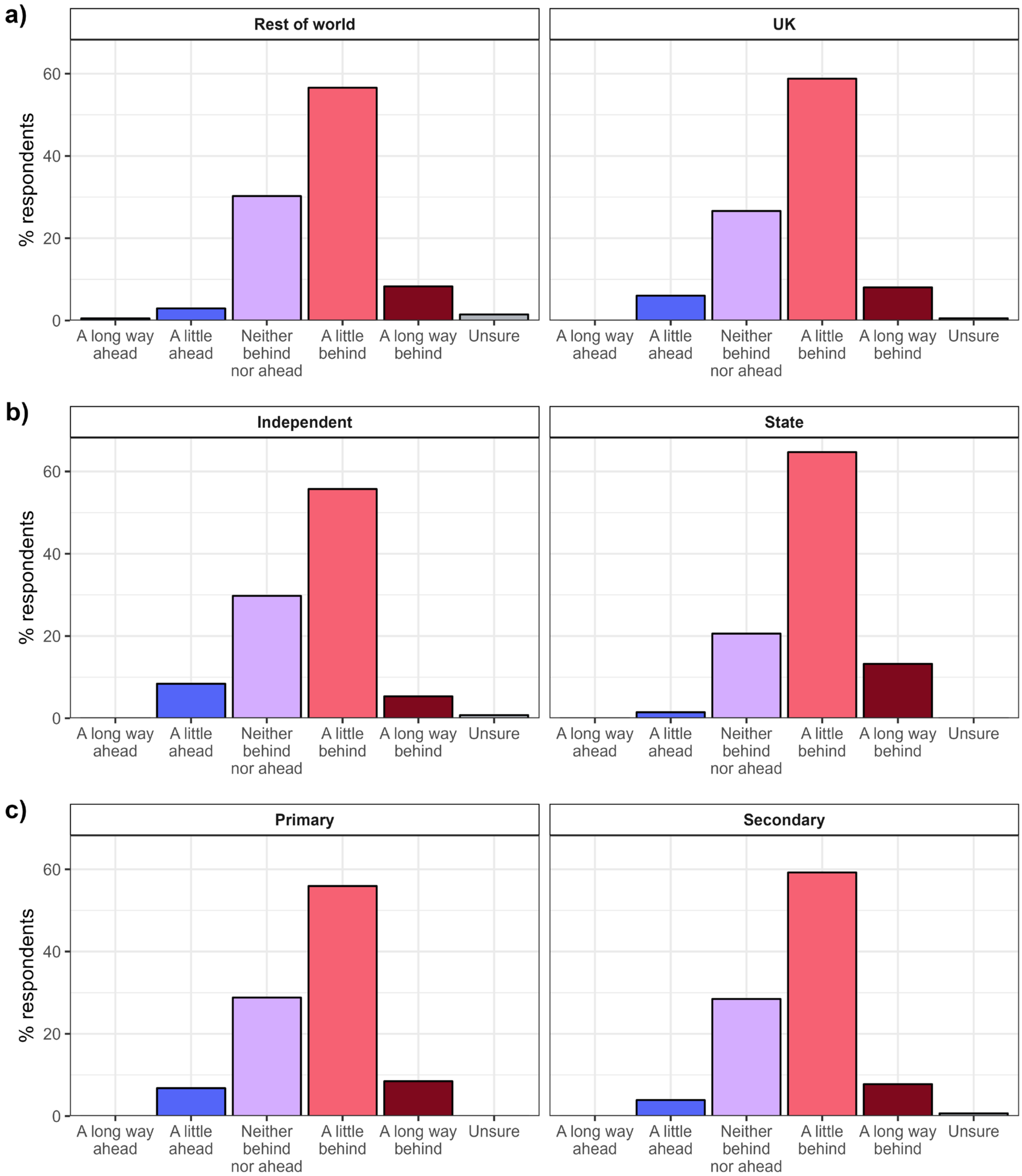

Figure 2: Responses to the question "How far ahead or behind in their curriculum learning do you feel most of your students are at the moment, compared to in a 'typical' year?", broken down into a) RoW and UK respondents, b) independent and state school respondents, and c) primary and secondary school respondents.

Overall, and in all comparison groups, the most common response was that students were "a little behind", with around 58 per cent of respondents overall saying this. Indeed, response patterns across all comparison groups were remarkably stable, with all showing broadly the same thing. One notable difference was, however, that estimates of students "a long way behind" were greater in state schools (13.2 per cent) than independent schools (5.3 per cent), as were estimates of students "a little behind" (state 64.7 per cent, independent 55.7 per cent). Note also, however, that the results show that a significant minority of respondents thought that their students were neither behind nor ahead, and a small minority thought they were ahead, showing that "loss" was not a universal experience.

To allow for comparisons with other estimates of learning loss, respondents were next asked "As a rough estimate, how far ahead or behind in their curriculum learning do you feel most of your students are at the moment?" For this, responses were analysed separately for those who thought their students were behind[1], and those who thought their students were ahead. Figures 3 and 4 show results for those who felt their students were behind; Appendix Table 2a gives counts and percentages for those who thought their students were behind, and Appendix Table 2b gives figures for those who thought they were ahead. Note that because so few respondents thought their students were ahead, these estimates are not plotted.

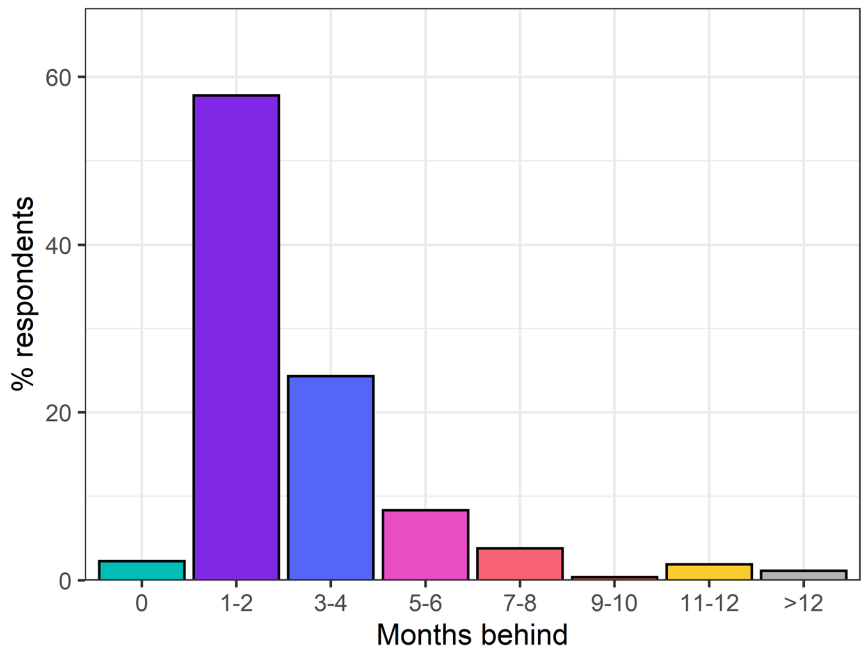

Figure 3: Overall estimates of how many months behind students were (considering only those 263 respondents who felt students were behind).

1 This splitting was carried out using responses to the question analysed in Figures 1 and 2, that is "How far ahead or behind in their curriculum learning do you feel most of your students are at the moment, compared to in a 'typical' year?" Anyone who answered "neither behind nor ahead" was not included in this step of the analysis.

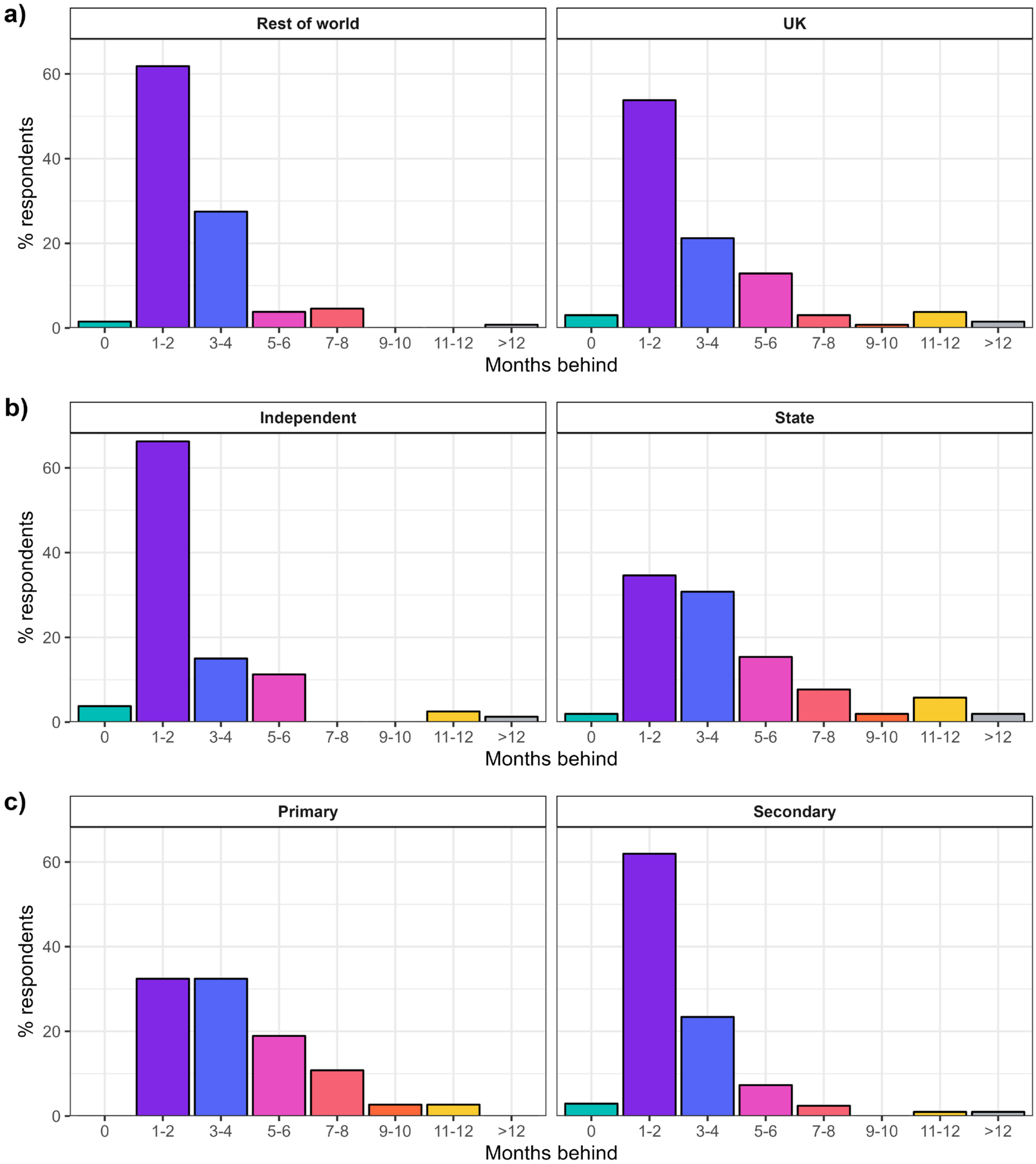

Figure 4: Estimates of how many months behind students were (considering only those respondents who felt students were behind), broken down into a) RoW and UK respondents, b) independent and state school respondents, and c) primary and secondary school respondents.

Considering estimates of students being behind, i.e., "learning loss", the most common response overall, and in most groups, was 1-2 months behind, and the next most common was 3-4 months behind (Figure 3, Table 2a). Note, however, that much larger estimates were not uncommon, with over 10 per cent of respondents overall giving an estimate of 5-6 months or greater. Some interesting contrasts emerged when looking at the comparison groups. First, estimates of loss were greater in state schools than independent schools, with 66 per cent of independent school respondents saying 1-2 months behind, compared to only 34.6 per cent in state schools; the remaining state school respondents contributed to higher response rates for all larger estimates of loss (e.g., 30.8 per cent for 3-4 months, 15.4 per cent for 5-6 months, 7.7 per cent for 7-8 months, etc.). Similarly, estimates of loss in primary schools were larger than those in secondary schools, with estimates of 1-2 months and 3-4 months behind equally common in primary school respondents, but with 62 per cent of secondary school respondents choosing 1-2 months behind.

Few respondents thought that their students were ahead (Appendix Table 2b). Of those that did think this, the most common response was 1-2 months ahead. The small numbers make it difficult to make robust comparisons between groups, but a notable observation is that only one state school respondent estimated their students to be ahead, compared to eleven independent school respondents.

Following these closed response questions, respondents were asked “If you feel your students are behind or ahead, in which aspects of the subject(s) that you teach are they behind or ahead (e.g., topics, skills)?” and a free text box was provided for answers. This was optional, but 289 responses were given. To provide a visual summary of responses, Figure 5 shows a word cloud of the most commonly used words. This indicates some key themes: skills, more than topics, had been lost, with core skills such as reading, writing, speaking and mathematics hit the hardest. Further, practical skills had been particularly affected by the shift to remote teaching. These were explored further when responses were read in full.

Figure 5: A word cloud showing the most common words used in responses to the question "If you feel your students are behind or ahead, in which aspects of the subject(s) that you teach are they behind or ahead (e.g., topics, skills)?" Larger words were used more frequently, and smaller words less frequently, with the font size proportional to the number of times the words were used.

Many responses were short, simply listing the areas where students were behind. For example, responses included "spelling, reading and maths", "speaking and writing skills", "writing and mathematics", "literacy and numeracy", "writing, reading, speaking and listening skills", and "handwriting has deteriorated [and] ability to write extended responses". These responses reflect a common theme: many of the areas lost were fundamental skills, particularly relating to maths and English. A number of respondents attributed this loss to remote learning, with comments like "their listening skills have suffered on account of the online teaching", and "students lack their calculating and critical thinking skills as most of the students either don't respond during the online sessions or they use digital devices to do the calculations, also someone might be helping them to answer the questions." Hence in spite of, or even because of, remote learning, many teachers felt fundamental skills had fallen behind.

Aside from literacy and numeracy, a commonly mentioned area of loss was practical skills, as practical sessions were difficult, if not impossible, to carry out remotely. Some comments mentioned practical skills in general, such as "behind with regards to practical skills", but several specific areas were also described. As may be expected, science was often mentioned, with comments such as "lab practical skills – experimental procedures, precautions, manipulating and interpreting data", and "our students have undertaken no proper practicals in science." A range of other subject areas were also mentioned, including physical education ("our younger students ... have missed the summer sports, so cricket, athletics and tennis"), art ("where they drag a bit is with the practical skills – ability to sketch quick and to sketch right, ability to work with dynamic compositions"), geography ("there has been no fieldwork, so the skills component has been seriously weakened"), music ("practical music skills – playing and composing"), and drama ("we haven't covered anything that has to do with the stage and the theatre space"). Loss, therefore, was not limited to areas easily monitored with standardised tests.

A further key area of loss was that of general study skills, of the type that may not be explicitly taught, but which are picked up from general schoolwork. Specific examples included "day-to-day management of workload/school habits", "acquisition of study habits", "soft skills and collaborative skills", "they definitely lack academic maturity", "social skills, communication and interaction", and "social skills and self regulation". One respondent noted that the loss of general skills could be particularly problematic for certain year groups: "Our year 7 … were remote for half of their year 6 and now are just back in school after 6 months remote in year 7. Their skills have really been impacted as has their loss of opportunity to 'grow' as secondary students or to have the leadership opportunities that would have come from being the top year in Primary." Hence, the range of skills considered "lost" was not limited to those explicitly taught or practised, but also included things that students gain simply from being at school.

A small number of respondents described areas where students were ahead. These most often related to areas where remote learning permitted extra focus or encouraged development of particular skills. Comments along these lines included "remote learning … allowed for more in-depth study of text", "definitely ahead in IT skills such as presenting and displaying data", and "they have increased their understanding of digital media such as photography and digital editing". Indeed, one respondent described opportunities presented by remote learning: "they have deeper understanding. Working remotely, we have been able to run seminar style lessons … This has led to much deeper understanding of content and concepts." Hence, although the majority of respondents described areas of loss, there were some areas where extra progress was possible in some cases.

The final emerging theme in free text comments related to variability in loss, with a reasonably large number of respondents describing variable impacts and suggesting reasons for this. Age appeared to have an impact, emphasised by comments such as "younger students … have been more adversely affected", and "the difference is most notable in the younger children who have had a significant proportion of their time in school disrupted." Student ability also appeared to play an important role, with comments such as "higher ability students are slightly ahead. Lower ability students are behind on exam technique, in-depth analysis and retrieval practice", "in general, lower attaining students found remote learning more challenging and some disengaged completely," and "those that need the most support with working in normal times have suffered the most." Indeed, one comment noted substantial individual-level variability: "every case is different – and some have thrived being left alone with more time, others have struggled with the lack of structure of remote learning". It appears, then, that learning loss was highly variable within and between groups.

This theme of variability of loss was picked up in the next survey question, which asked respondents "How much has the educational gap between your most able and your least able students changed since the start of the pandemic?" Results are presented in Figures 6 and 7, and in Appendix Table 3. The biggest response category overall was that gaps had "increased a little" (42.8 per cent), followed by gaps having "increased a lot" (25.2 per cent), meaning that 68 per cent of respondents thought gaps had increased. Note, however, that a significant minority (9.4 per cent overall) thought that gaps had actually decreased.

Although "increased a little" was the biggest category in each comparison group, some differences were still evident: estimates of increased gaps (i.e., "a little" and "a lot" combined) were more common in the UK (76.4 per cent) than in RoW (60.0 per cent), more common in state schools (86.8 per cent) than in independent schools (71.0 per cent), and more common in secondary schools (72.1 per cent) than in primary schools (59.3 per cent). Perhaps the most notable difference was between state schools and independent schools: 36.8 per cent of respondents in state schools felt that gaps had "increased a lot", compared to 17.6 per cent in independent schools. Although there was no specific free text question about the size of educational gaps, responses to other questions (not analysed in detail here) described variability in access to technology, parental support, and engagement, as well as effects of ability level. Therefore, a range of factors could affect the extent of learning loss and, within a group of students facing varying circumstances, influence the resulting educational gaps.

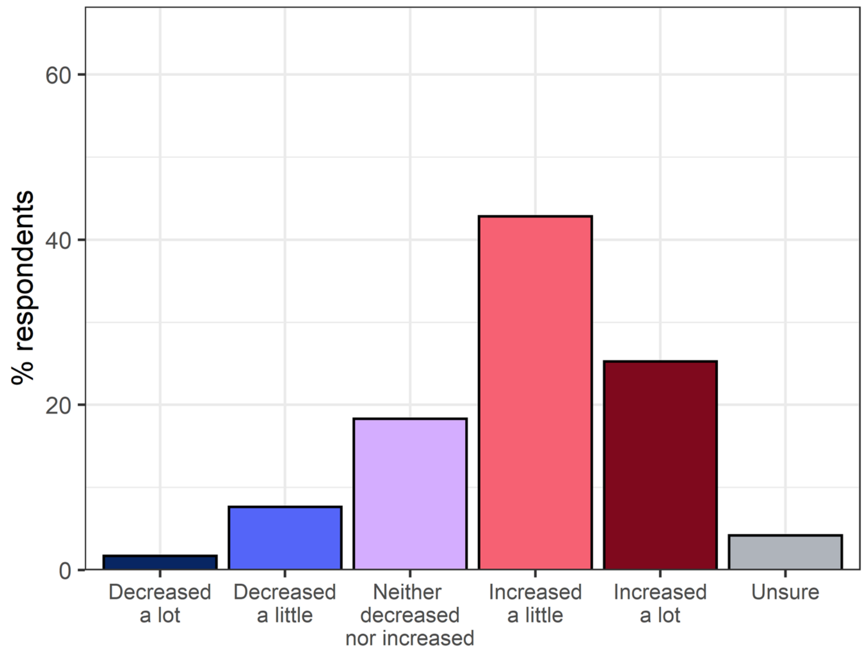

Figure 6: Overall responses to the question "How much has the educational gap between your most able and your least able students changed since the start of the pandemic?"

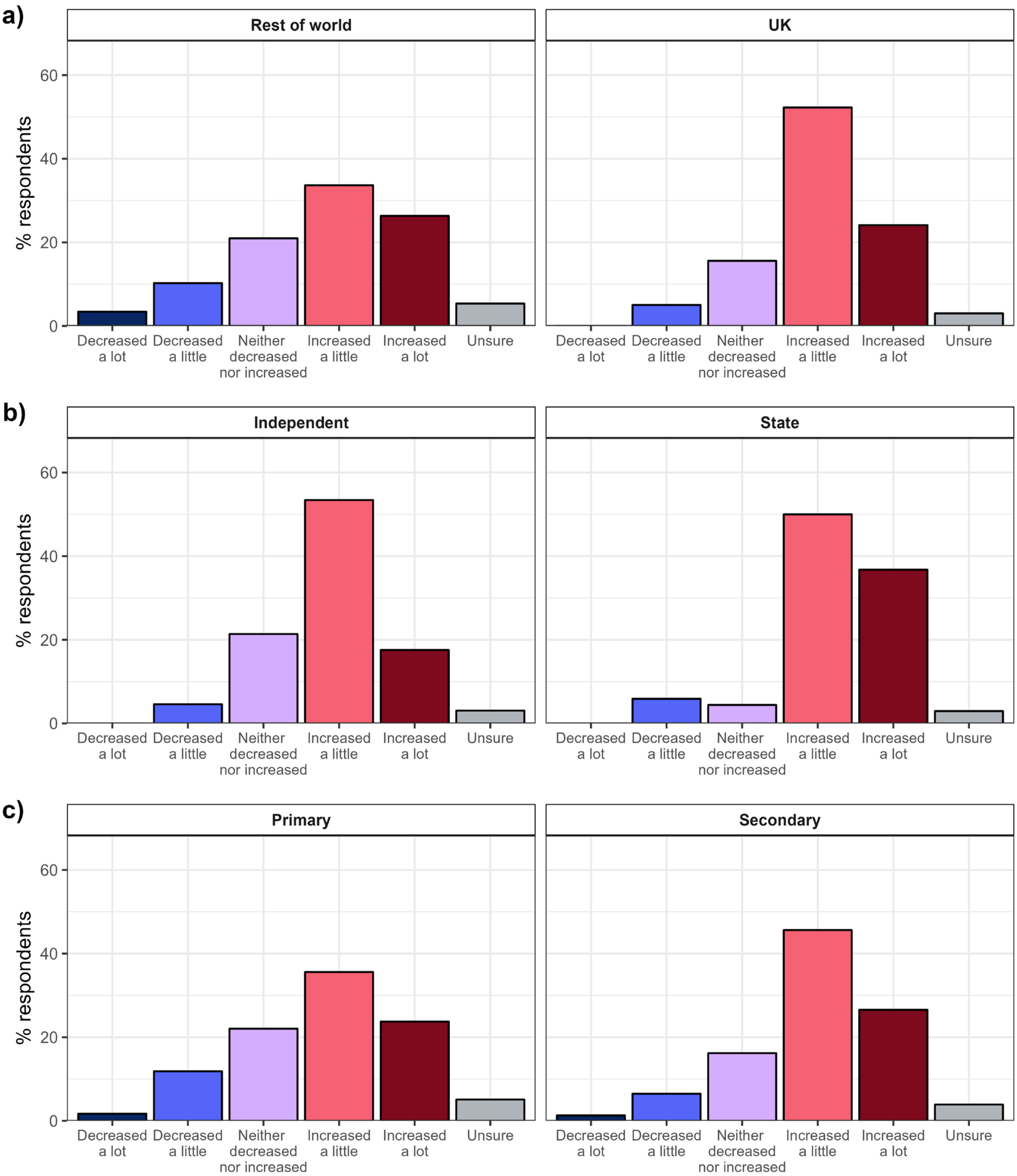

Figure 7: Responses to the question "How much has the educational gap between your most able and your least able students changed since the start of the pandemic?" broken down into a) RoW and UK respondents, b) independent and state school respondents, and c) primary and secondary school respondents.

Discussion

Learning loss is frequently discussed as a major consequence of the disruption to education during the Covid-19 pandemic. Here, responses to a survey sent to teachers after one year of teaching through the pandemic have helped us to understand more about how much, and what, was "lost".

Before exploring the results in detail, it is worth considering the limitations of the study. Perhaps the largest limitation is that the sample of respondents is relatively small. Although over 400 responses were received, spread across 198 schools, this is a tiny fraction of the number of teachers and schools in the world. Further, the sample composition is not representative of the actual composition of schools and teachers, either within the UK or the wider world. This means that the findings might over-emphasise particular experiences and under-emphasise others. The subgroup comparisons allow at least some of the effects of this to be explored, as the influence of key sources of variation could be examined, but even these comparisons cannot claim to be fully representative of the groupings considered. Hence, while the results can tell us valuable things about teachers' experiences of learning loss during the Covid-19 pandemic, we cannot tell the extent to which they capture the full range of views. Nevertheless, the sample is large enough, and responses detailed enough, that we can still draw conclusions and make inferences from the results.

A main finding worth emphasising is that a majority of teachers did feel that their students were, on average, behind where they would be in a typical year. That is, the phenomenon of "learning loss" does seem to have occurred. However, a large minority of respondents did not observe an overall loss, and a small minority found that some students were ahead compared to a typical year. Therefore, these results support the idea that the disruption to education caused students to fall behind, but it does not appear to be a truly universal experience, despite the global nature of the pandemic.

Intriguingly, estimates of the amount of learning lost were similar to those calculated via more quantitative studies, which have typically indicated loss of 1-2 months, albeit with much larger estimates in some cases. The most common estimate here was 1-2 months behind, but 3-4 months behind was also a common response. This suggests that teachers' perceptions of lost progress are fairly accurate, and in turn provides a degree of support for findings from quantitative studies (e.g., those reported by Newton, 2021). It is notable, however, that some much larger estimates were made, with over 10 per cent of those respondents who thought their students were behind estimating 5 months or greater. Bearing in mind that these were estimates of *average* loss, it raises concerns that some groups may have been very strongly affected. Most estimates were, however, somewhat reassuring: a "loss" of 1-2 months or even 3-4 months is less than the length of the disruption, which suggests that teachers, schools, and whole education systems, managed to counteract at least some of the possible negative impacts.

The findings discussed thus far point to what is, perhaps, a greater concern than the presence of "loss" itself. That is, the impacts of the disruption were variable and unevenly distributed. As noted above, loss was not universally experienced, and even among those who were considered to be behind, the extent of loss varied. Free text comments suggested that there could be strongly varying impacts within the same class or school, with individual family circumstances and student personalities influencing outcomes. This variability was further emphasised by comparisons between subgroups of respondents, which indicated some structural aspects to learning loss. That is, students in state schools and younger students appeared to have experienced greater impacts than those in independent schools and older students respectively. Similar patterns have been noted elsewhere (e.g., Howard, Khan & Lockyer, 2021; Major, Eyles & Machin, 2020; Open Data Institute, 2020), again reinforcing the emerging picture of important variability in loss.

Our study also allowed us to consider the nature of what had been lost. Other attempts to examine loss often focus on standardised tests of numeracy and language skills, so can only really draw conclusions about these areas. Here, by asking teachers about what they felt had been lost (or, indeed, gained), we were able to look beyond these core areas. Many comments reflected on the loss of fundamental skills, such as writing and reading. Although remote learning would have clearly included the use of such skills, they appear not to have developed in the same way during that period. Some skills could not be covered remotely, notably practical science, but also practical aspects of sports, music and drama; again, these were all mentioned as areas where loss had occurred. Further, comments discussed the loss of more general skills, such as communication, workload management and social skills; such skills are not always formally taught, but develop as part of school life. Therefore, results indicate that "learning loss" appears to not be the uniform loss of all learning, but instead reflects the loss, or lack of development, of particular skills.

The above discussion of the nature of learning loss has important implications not just for the way we understand it, but also how we respond to it. If learning loss was uniform in both extent and nature, catching up could be achieved simply by providing extra hours of teaching, covering what was missed. However, the variability means that some students will need much more support, while others, who may have progressed *more* than in a normal year, may not need any support. The structural elements of loss identified, including variation between age groups and school types, also introduce an equality angle to the discussion: whole groups of students have been affected more than others, meaning that existing inequalities have widened, and bringing into question whether certain groups need focused support. Moreover, it seems feasible to provide specific catch-up time on some areas (e.g., numeracy or practical science skills), but other areas of loss may be better served by supporting the transition back to normal schooling (e.g., general study and social skills). Therefore, responses to learning loss must consider who needs support, what needs to be covered, and whether the loss would be recovered naturally over time anyway. Despite the well-intentioned focus on rapidly responding to impacts of the pandemic, responses should be carefully considered to ensure efficient and equitable use of catch-up resources.

Indeed, one author has characterised the challenges of responding to learning loss as a "trap", in which a strong focus on quantifying loss, a lack of acknowledgement of variability, and a focus on numeracy and language all lead to inefficient or ineffective responses (Zhao, 2021). A particular risk the author raises is that focus on "catching up" in numeracy and literacy draws resources away from other areas. Instead, the author argues that responses should use teachers' professional judgement to identify the extent of support required, and consider a wide range of educational outcomes. Moreover, the author points out that there may be opportunities: the increased engagement of families, increased use of independent learning, and innovations introduced by remote learning can all be developed in the post-pandemic world. Other authors have made similar claims: by developing effective catch-up approaches there may be an opportunity to "build back better" (Kaffenberger, 2021). To "build back better" and avoid "traps", it is important to consider what has been lost, by whom, and what will best help them to catch up.

The research presented here does not challenge the narrative of learning loss: a majority of survey respondents from around the world, and from different school types, reported that their students were behind where they would be in a normal year. The estimates of loss experienced are similar to those from other studies using entirely different methods, adding credence to the findings here. However, the findings discussed above help to qualify what loss means: it is variable within and between groups, with some students even showing *extra* progress, and it seems to have impacted development of skills more than coverage of curriculum content. Efforts to make up for learning loss are well intentioned and important, especially given some of the large estimates of loss reported here and the likely impacts on educational equality. But the findings here caution us not to look for simple, large-scale fixes, which could increase pressure on teachers and students, and which may not be applied efficiently. Instead, consideration of individual needs and circumstances, use of teachers' professional judgement – and even consideration of new opportunities and what may have been gained – could ensure that the right kind of support is provided to those who need it, in turn helping to mitigate some of the longer-term impacts of the disruption.

Acknowledgements

We are grateful to our CEM colleagues Irenka Suto and Chris Jellis for support in developing and distributing the survey. We are grateful to all colleagues who provided feedback and guidance during survey development. We thank all participants for giving their time to complete the survey and contributing to the research.

References

Chen, L.-K., Dorn, E., Sarakatsannis, J., & Wiesinger, A. (2021). *Teacher survey: Learning loss is global – and significant*. McKinsey & Company.

Donnelly, R., & Patrinos, H. A. (2021). Learning loss during Covid-19: An early systematic review. *Prospects*. https://doi.org/10.1007/s11125-021-09582-6

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2021). *COVID-19 and education: The lingering effects of unfinished learning.* McKinsey & Company.

Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *PNAS, 118*(17), e2022376118. https://doi.org/10.1073/pnas.2022376118

Howard, E., Khan, A., & Lockyer, C. (2021). *Learning during the pandemic: review of research from England.* Ofqual, Coventry, UK.

Kaffenberger, M. (2021). Modelling the long-run learning impact of the Covid-19 learning shock: Actions to (more than) mitigate loss. *International Journal of Educational Development*, 102326. https://doi.org/10.1016/j.ijedudev.2020.102326

König, C., & Frey, A. (2022). The Impact of COVID-19-Related School Closures on Student Achievement – A Meta-Analysis. *Educational Measurement: Issues and Practice, 41*, 16–22. https://doi.org/10.1111/emip.12495

Major, L. E., Eyles, A., & Machin, S. (2020). *Generation COVID: Emerging work and education inequalities*. Centre for Economic Performance, London, UK. https://cep.lse.ac.uk/pubs/download/cepcovid-19-011.pdf

Newton, P. (2021). *Learning during the pandemic: quantifying lost learning. Report 3 of 5 on learning during the 2020 coronavirus (COVID-19) pandemic.* Ofqual, Coventry, UK.

Open Data Institute. (2020). *Data about children's lives during the pandemic*. Open Data Institute, London, UK. https://theodi.org/article/data-about-children-during-covid19/

Renaissance Learning & the Education Policy Institute. (2021). *Understanding Progress in the 2020/21 Academic Year. Findings from the summer term and summary of all previous findings.* Department for Education, London, UK.

Rose, S., Twist, L., Lord, P., Rutt, S., Badr, K., Hope, C., & Styles, B. (2021). *Impact of school closures and subsequent support strategies on attainment and socio-emotional wellbeing in Key Stage 1: Interim Paper 1.* Education Endowment Foundation, London, UK.

Sharp, C., Nelson, J., Lucas, M., Julius, J., McCrone, T., & Sims, D. (2020). *Schools' responses to Covid-19: The challenges facing schools and pupils in September 2020*. National Foundation for Educational Research, Slough, UK.

Twist, L., Jones, E., & Treleaven, O. (2022). *The Impact of Covid-19 on pupil attainment*. National Foundation for Educational Research, Slough, UK.

Zhao, Y. (2021). Build back better: Avoid the learning loss trap. *Prospects*. https://doi.org/10.1007/s11125-021-09544-y

Appendix

This appendix contains tables that report the underlying counts and percentages used to create the figures in the main article.

Table 1: Raw counts and percentages for responses to the question "How far ahead or behind in their curriculum learning do you feel most of your students are at the moment, compared to in a 'typical' year?"

		A long way ahead	A little ahead	Neither behind nor ahead	A little behind	A long way behind	Unsure
Overall	N	1	18	115	233	33	4
	%	0.2 %	4.5 %	28.5 %	57.7 %	8.2 %	1.0 %
RoW	N	1	6	62	116	17	3
	%	0.5 %	2.9 %	30.2 %	56.6 %	8.3 %	1.5 %
UK	N	0	12	53	117	16	1
	%	0.0 %	6.0 %	26.6 %	58.8 %	8.0 %	0.5 %
Independent	N	0	11	39	73	7	1
	%	0.0 %	8.4 %	29.8 %	55.7 %	5.3 %	0.8 %
State	N	0	1	14	44	9	0
	%	0.0 %	1.5 %	20.6 %	64.7 %	13.2 %	0.0 %
Primary	N	0	4	17	33	5	0
	%	0.0 %	6.8 %	28.8 %	55.9 %	8.5 %	0.0 %
Secondary	N	0	12	88	183	24	2
	%	0.0 %	3.9 %	28.5 %	59.2 %	7.8 %	0.6 %

Table 2: Raw counts and percentages for estimates of how far behind or ahead students were, for a) respondents who thought their students were ahead, and b) respondents who thought their students were behind.

a) Responses to "I estimate that my students were behind by..."

		0 months	1-2 months	3-4 months	5-6 months	7-8 months	9-10 months	11-12 months	Over 12 months
Overall	N	6	152	64	22	10	1	5	3
	%	2.3 %	57.8 %	24.3 %	8.4 %	3.8 %	0.4 %	1.9 %	1.1 %
RoW	N	2	81	36	5	6	0	0	1
	%	1.5 %	61.8 %	27.5 %	3.8 %	4.6 %	0.0 %	0.0 %	0.8 %
UK	N	4	71	28	17	4	1	5	2
	%	3.0 %	53.8 %	21.2 %	12.9 %	3.0 %	0.8 %	3.8 %	1.5 %
Independent	N	3	53	12	9	0	0	2	1
	%	3.8 %	66.2 %	15.0 %	11.2 %	0.0 %	0.0 %	2.5 %	1.3 %
State	N	1	18	16	8	4	1	3	1
	%	1.9 %	34.6 %	30.8 %	15.4 %	7.7 %	1.9 %	5.8 %	1.9 %
Primary	N	0	12	12	7	4	1	1	0
	%	0.0 %	32.4 %	32.4 %	18.9 %	10.8 %	2.7 %	2.7 %	0.0 %
Secondary	N	6	127	48	15	5	0	2	2
	%	2.9 %	62.0 %	23.4 %	7.3 %	2.4 %	0.0 %	1.0 %	1.0 %

b) Responses to "I estimate that my students were ahead by..."

		0 months	1-2 months	3-4 months
Overall	N	2	14	3
	%	10.5 %	73.7 %	15.8 %
RoW	N	2	4	1
	%	28.6 %	57.1 %	14.3 %
UK	N	0	10	2
	%	0.0 %	83.3 %	16.7 %
Independent	N	0	9	2
	%	0.0 %	81.8 %	18.2 %
State	N	0	1	0
	%	0.0 %	100.0 %	0.0 %
Primary	N	0	3	1
	%	0.0 %	75.0 %	25.0 %
Secondary	N	2	8	2
	%	16.7 %	66.7 %	16.7 %

Table 3: Raw counts and percentages for responses to the question "how much has the educational gap between your most able and your least able students changed since the start of the pandemic?"

		Decreased a lot	Decreased a little	Neither	Increased a little	Increased a lot	Unsure
Overall	N	7	31	74	173	102	17
	%	1.7 %	7.7 %	18.3 %	42.8 %	25.2 %	4.2 %
RoW	N	7	21	43	69	54	11
	%	3.4 %	10.2 %	21.0 %	33.7 %	26.3 %	5.4 %
UK	N	0	10	31	104	48	6
	%	0.0 %	5.0 %	15.6 %	52.3 %	24.1 %	3.0 %
Independent	N	0	6	28	70	23	4
	%	0.0 %	4.6 %	21.4 %	53.4 %	17.6 %	3.1 %
State	N	0	4	3	34	25	2
	%	0.0 %	5.9 %	4.4 %	50.0 %	36.8 %	2.9 %
Primary	N	1	7	13	21	14	3
	%	1.7 %	11.9 %	22.0 %	35.6 %	23.7 %	5.1 %
Secondary	N	4	20	50	141	82	12
	%	1.3 %	6.5 %	16.2 %	45.6 %	26.5 %	3.9 %

Which assessment is harder? Some limits of statistical linking

Tom Benton and Joanna Williamson (Research Division)

Introduction

Equating methods are statistical processes whose purpose is to put scores from different assessments onto the same scale. A key application of equating is to determine equivalent scores when candidates for the same qualification can take alternate versions of certain assessment components. For example, candidates who are in different time zones or who take the same qualification at a different time of year may sit different versions of a written examination component, and it is necessary to know which scores represent the same level of achievement on the different assessment versions so that no candidate is disadvantaged. Definitions of equating stress that equating is for adjusting between alternate versions of assessments targeting the same content at the same level, with the aim that scores from the different versions can be used "interchangeably" (Kolen & Brennan, 2014, p. 2).

The statistical processes used in equating have also, however, been extended to compare pairs of assessments that do not meet these strict criteria. There is often great interest in the comparability of assessment scores from related assessments targeting the same construct at different levels, from parallel qualifications targeting the same subject at the same level, and from assessments of the same qualification type that assess different subjects. The use of equating methods and close variants to statistically "link" assessments for such comparisons has a different conceptual basis to equating in the clear sense that statistical adjustments cannot make the scores from a Physics exam and a History exam "interchangeable". There are however high-stakes situations in which such scores are in fact interpreted interchangeably (e.g., school league tables, or a university place conditional on achieving three A Levels at grades AAB), providing ample motivation for asking whether certain assessments are "too hard" or "too easy" in comparison with others. Despite careful debate over the basis for statistical linking and the precise conclusions that can and cannot be drawn (e.g., Mislevy, 1992, pp. 21-26; Newton, 2010), including in the literature on inter-subject comparability (e.g., Bramley, 2011; Coe, 2008; Newton, 2012), it can be tempting to apply equating methods and conclude that they have provided a definitive answer regarding whether a qualification is harder or easier than others.

The purpose of this article is to explore how accurately various equating methods are able to equate between identical assessments. It offers a novel demonstration of some limits of statistical equating by making use of pairs of live assessments that are "cover sheet" versions of each other, that is, identical assessments with different assessment codes. Such pairs occur most commonly where the same assessment is a component of corresponding qualifications of different types (e.g., an IGCSE and O Level in the same subject) or a component of related qualifications of the same type (e.g., IGCSE Combined Science and IGCSE Co-ordinated Sciences). The fact that the assignment of students to particular cover sheet versions is a non-random process means that this context may provide a more realistic evaluation of various equating techniques than others. In particular, the equating methods will have to address issues of differences in the abilities and subject choices of candidates taking different qualifications that occur in real practical situations. At the same time, the evaluation of the equating methods' accuracy is made straightforward by the fact that the true equating relationship is known: since the two assessments in a cover sheet pair are identical, the scores from the two assessments are already on the same scale, and the true equating relationship is the one that maps each score to itself (in mathematical terms, the identity function).

How to link assessments

The outcome of equating two assessments is a statistical transformation or equating function that allows scores from one assessment to be interpreted on the same scale as scores from the second assessment. Two things are needed to generate the statistical transformation or equating function: firstly, locating or collecting some data that links candidate performance on the two assessments, and secondly, making a decision about which definition of "same standard" the equating function should preserve.

In some equating designs, information linking candidate performance on the two assessments is obtained directly, by having a single group of representative candidates take both assessments. Alternatively, candidates may be randomly assigned to sit one or other of the two assessments, and for sufficiently large groups, the groups can be assumed equivalent. In these designs, differences in performance can be interpreted as representing differences in the difficulty of the two assessments, rather than differences in the candidates sitting them.

In high-stakes live assessments such as IGCSEs, O Levels and A Levels, security concerns prevent the pre-testing of assessments, and it is not possible to assign candidates randomly to different live papers. Where the groups sitting each assessment cannot be assumed equivalent, it is – clearly – more challenging to judge comparable standards in the two assessments, as this has to be disentangled from differences in the ability of the two groups. Equating designs for non-equivalent groups require some link between the assessments of interest.

Where available, this link can be achieved via a subset of common items that feature on both assessments (as seen, for example, in tiered GCSE Mathematics). The fact that these items are taken by both candidate groups allows them to function as a reference point or "anchor" for understanding the group differences (hence Non-Equivalent Groups with Anchor Test or NEAT equating design). For many pairs of assessments, however, there is no subset of items forming an internal anchor test, and an external link or anchor must be found. For GCSEs and A Levels in England, prior attainment, at Key Stage 2 and GCSE respectively, is typically used in place of an anchor (see Bramley & Vidal Rodeiro, 2014). In other scenarios, particularly where prior attainment is unavailable, an external link might be identified from common components, that is, assessments taken by candidates from both groups alongside the assessments being equated.

Previous work by Benton (2017) developed a method for going beyond co-components and taking into account all the information linking candidates. The core idea is a summary measure known as the ISAWG (Instant Summary of Achievement Without Grades), a measure of ability that summarises each candidate's performance across multiple assessments on a single scale, whichever assessments they have taken. The ISAWG value for each candidate can be defined informally as "the single number that most accurately reflects the standardised marks they have achieved on whichever assessments they have taken" (Benton, 2017, p. 6). When used to equate between assessments, the ISAWG measure therefore incorporates information about candidates' performance on all co-components (if any exist), but also assessments that are not co-components[1].

An important theoretical objection to equating assessments via co-components or ISAWG measures is the defensibility of comparing assessments in one subject using data from assessments designed to measure candidates' abilities in different subjects or qualifications. Besides assessing different content, factors that can undermine comparisons include differences in teaching and levels of student motivation, and whether an assessment is compulsory or the result of student choice. Data on candidates' achievement in different assessments can be used in such a way that the most relevant information is prioritised over less relevant information (e.g., by restriction to related subjects, or to similar qualification types, or prioritisation of co-components according to correlation with assessment scores and candidate numbers), but the concern is a valid one, and has been extensively debated. The theoretical basis for pursuing a measure such as ISAWG is Spearman's (1904) theory of general ability or "g", which would suggest that "although different tests may measure slightly different skills, all of them should relate to each candidate's 'fundamental function' (or 'g')" (Benton,

1 The technical procedure for calculating ISAWG is equivalent to carrying out Principal Components Analysis on a data set including all of the assessments offered by Cambridge International and OCR in a single session – with missing values included, since no candidate takes all available assessments – and taking the first principal component for each candidate. Although other research has also investigated how to incorporate information from covariates into equating (e.g., Andersson et al., 2013; Wiberg & Branberg, 2015), the ISAWG is uniquely well suited for equating using very large sets of covariates (in this case, assessments) with highly variable missing data patterns (see Benton, 2017, p. 8).

2017, p. 6). This, in turn, should provide a reasonable basis for estimating the candidate's likely achievement on other assessments. In the context of setting grade boundaries or cut scores for standard maintaining, previous work has shown that co-component and ISAWG equating methods are promising (Benton, 2017). Importantly, in the context of setting grade boundaries, equating outcomes can be and are considered alongside multiple other sources of evidence, including expert judgements about question papers and candidate scripts, and sometimes alternative types of statistical evidence.

Once data linking the assessments to be equated has been identified, there are multiple ways to define an equating relationship. One widely used approach is equipercentile equating, in which scores from Test X and Test Y are considered equivalent if they represent the same percentile rank for the specified population. Equipercentile equating allows for non-linear relationships between Test X and Test Y: for example, the equating function may indicate that Test X is easier than Test Y at the very top and bottom of the score range, but not in the middle of the range. Equipercentile equating requires more data than some other methods, but is less restrictive in its assumptions and requirements and hence is suitable for the type of assessments considered in this article, where lack of data is not a problem.

Equating percentile ranks for the complete and non-equivalent groups taking Test X and Test Y would of course not account for any differences in group ability, and there are two main approaches to dealing with this. In frequency estimation (FE) equipercentile equating, the candidate groups for Test X and Test Y are first weighted so that they are equivalent in terms of their anchor test score distributions (i.e., to create equivalent groups, so far as we are able, for which we have both Test X and Test Y scores). Using the weighted data, the score distributions for Test X and Test Y are created. These are then used to equate percentile ranks. As an alternative to FE, the chained method equates percentiles first from Test X to the anchor test within the Test X candidates, then equates percentiles from the anchor test to Test Y within the Test Y candidates.

Method

Several equating methods were investigated by exploring how accurately they equated between pairs of identical assessments with different cover sheets. Each method produced an estimated equating function linking the scores between the two assessments in each pair (Test X and Test Y). These equating functions were evaluated by comparison to the true equating function, which, for all pairs, was the identity function. For each Test X score, the difference between the estimated equated score and actual equivalent score (equal to the Test X score) was calculated. The differences between the estimated equated scores and actual equivalent scores were then summarised in terms of the cumulative percentage of candidates achieving each score or above. This information indicated the differences in pass rates that would result from cut scores at any chosen point, allowing the equating errors to be interpreted in terms of their impact rather than just magnitude.
Five equating techniques were investigated: four versions of equipercentile

equating with an anchor test or measure, and, as a contrasting but widely used approach, Rasch equating[2]. To support fair comparison between the equating methods, each method was restricted to considering the same set of possible co-components. For each cover sheet pair, the set of usable co-components was defined to be the 20 largest components (by joint *N* also taking Tests X and Y) taken by at least 100 and at least 5 per cent of candidates taking Test X, and also at least 100 and at least 5 per cent of the candidates taking Test Y.

The details of the five methods were as follows:

1. **Single co-component (FE)**

Components were equated by choosing a single co-component to use as an anchor test, in frequency estimation (weighted) equipercentile equating. The co-component chosen from the set of (up to) 20 usable co-components was that with the highest minimum correlation with Test X and Test Y. Single co-component equating was investigated due to its simplicity as well as good performance in prior equating studies (Benton, 2017).

2. **Single co-component (chained)**

Components were equated using a single co-component as an anchor, in chained equipercentile equating. As in Method 1, the co-component selected was that which had the highest minimum correlation with Test X and Test Y. It was considered important to test chained as well as frequency estimation methods, since frequency estimation is recommended only when groups are "reasonably similar" (Kolen & Brennan, 2014, p. 146), and there is evidence that chained methods may be more successful when the abilities of Test X and Test Y candidates in fact differ meaningfully (Benton, 2017).

3. **ISAWG (FE)**

Components were equated using candidates' ISAWG measure in place of an anchor test score in frequency estimation equipercentile equating. The ISAWG measure was recoded into integers 0-19 before equating was carried out (since nearly all equating methods expect integer anchor test scores).

4. **ISAWG (chained)**

Components were again equated using the separately calibrated ISAWG measure in place of an anchor test, but within chained equipercentile equating. In contrast to the standard ISAWG measure (Benton, 2017), the ISAWG measures used in Methods 3 and 4 were separately calibrated for each cover sheet pair. The calculation of ISAWG values was restricted to using scores from that pair's set of (up to) 20 usable co-components as well as the two components themselves. Note that the two components in the pair being equated were *not* themselves treated as cover sheet versions of each other. That is, the ISAWG calculation considered these two components as entirely separate assessments, while, if they

2 In contrast to the four equipercentile equating procedures, which are observed-score equating methods, the Rasch equating method is a form of item-response theory (IRT) true-score equating in which scores from Test X and Test Y are considered equivalent when they correspond to the same level of underlying ability construct (see Kolen & Brennan, 2014, pp. 175, 213).

were among the relevant co-components, all other cover sheet pair equivalences remained "known" to the ISAWG calculation.

5. Rasch equating

For each cover sheet pair, Test X, Test Y and the corresponding (up to) 20 usable co-components were first analysed as a (up to) 22 "item" test using a polytomous extension of the Rasch model (an Extended Nominal Response Model fitted in R, using the package Dexter (Maris et al., 2021)). This allowed raw scores from Test X and Test Y to be related to a single unidimensional ability scale. Scores on Test X and Test Y were then linked so that for each Test X score, the equated Test Y score was the Test Y score corresponding to the same point on the ability scale.

Data set and description of equating context

Equating was carried out on pairs of IGCSE and O Level components taken by Cambridge International candidates in summer 2018. Each pair consisted of identical assessments with different cover sheets, taken by non-overlapping groups of candidates. Analysis was restricted to pairs with at least one co-component suitable for equating, defined to be an assessment component taken by at least 100 and at least 5 per cent of the Test X candidates, and by at least 100 and at least 5 per cent of the Test Y candidates. Analysis was further restricted to pairs in which each assessment was taken by at least 3000 candidates who also took at least one co-component, and where the total available marks were at least 50, to avoid the results reflecting the difficulties of equating with too few candidates or too few marks. Individual qualifications were included only once. Where multiple cover sheet pairs from the same qualification met the conditions for inclusion, the pair with the higher number of available marks was retained, and if multiple pairs still remained, the pair of components with the smallest difference in raw mark means was retained.

Eight pairs of assessments met the above conditions, covering subjects from English as a Second Language (ESL) to Mathematics (Table 1). All the assessments were externally assessed written examinations. The first pair of ESL components (ESL 1) assessed both reading and writing, while the second pair (ESL 2) assessed writing only. Both the Maths 1 and Maths 2 component pairs consisted of two-hour written tests, but belonged to different mathematics qualifications. All pairs had at least five usable co-components. Where candidates are non-randomly assigned to assessment versions, the candidate groups taking each version may differ substantially – particularly if taught in different school systems – and the differences in mean assessment scores shown in Table 1 reflect this. For some pairs the mean marks achieved in Test X and Test Y were extremely close, for others there was a moderate difference, and for two pairs the difference was very large.

Table 1: Description of component pairs investigated.

Component pair	Max. mark	Test X N	Test Y N	Number of usable co-components (capped at 20)	Difference between Test X and Test Y mean scores (as per cent of maximum mark)
Business	80	5375	10 788	7	0.87
Computing	75	7627	4222	9	2.35
Economics	90	5898	9469	7	4.14
ESL 1 (Reading and Writing)	90	5906	9031	20	0.97
ESL 2 (Writing only)	60	3103	24 889	5	12.93
History	60	3088	12 033	18	6.57
Maths 1	80	4726	6263	7	5.28
Maths 2	104	3518	8104	10	20.55

As noted in the description of equating methods, the co-component selected for single co-component equating (Methods 1 & 2) was that (from the set of usable co-components) which had the highest minimum correlation with Test X and Test Y scores. Table 2 shows that these minimum correlations were generally high. The pairs for which the single best co-component had the lowest correlations with scores were History (0.52 with Test X) and ESL 2 (0.62 with Test Y). For History, the single chosen co-component was also only taken by a relatively small minority of those taking Tests X and Y. For Business, Computing, Economics and the two Maths pairs, correlations between the single best co-component and component scores were all 0.8 or higher. Correlations between component scores and the separately calibrated ISAWG measures used in Methods 3 and 4 were also high. The lowest correlation occurred for ESL 2 (0.72), and for Business, Computing, Economics and the two Maths pairs, correlations between the ISAWG measures and component scores were again particularly high (around 0.9). Part of the reason for these high correlations is, of course, that the components themselves (that is, Tests X and Y) contribute to the calculation. The use of multiple co-components also avoids potential loss of data by restricting to a single co-component.

Table 2: Correlation of component scores with single co-component anchor measures and ISAWG. Number of students with available score on the single co-component are also shown.

Component pair	Method 1 & 2 co-component with Test X		Method 1 & 2 co-component with Test Y		Correlation of ISAWG with Test X, Test Y	
	Correlation	N	Correlation	N	Test X	Test Y
Business	0.80	5373	0.82	10 774	0.91	0.92
Computing	0.80	7619	0.81	4221	0.88	0.91
Economics	0.80	5885	0.80	9457	0.93	0.93
ESL 1	0.73	5880	0.73	9004	0.83	0.82
ESL 2	0.62	3100	0.69	24 871	0.88	0.72
History	0.52	337	0.62	1500	0.82	0.83
Maths 1	0.91	4726	0.89	6262	0.94	0.92
Maths 2	0.88	3518	0.88	8089	0.90	0.96

A first indication of the differences between Test X and Test Y candidate groups was given by the differences in component scores reported in Table 1. To allow a closer look at any between-group differences, Figure 1 compares the standardised difference between Test X and Test Y candidates' component scores[3] (on the x axis) with the standardised difference in their ISAWG measures[4] (on the y axis). Figure 1 demonstrates, firstly, that there was a high level of agreement between the two measures in terms of which candidate group was higher performing. Both the direction and size of the ability difference indicated by the standardised score difference was generally reflected by the standardised ISAWG difference. The largest discrepancy was for ESL 2, where component scores indicated a standardised difference of -1 between candidate groups, whereas the ISAWG measure indicated a difference of -0.67 standard deviations. Secondly, Figure 1 highlights that for some of the cover sheet pairs, the difference between Test X and Test Y candidate groups was rather large, confirming that this real-world equating context included a high level of challenge. As a rule of thumb, Kolen and Brennan (2014, p. 301) note that equating can be "especially troublesome" where group differences are larger than 0.5 standard deviations. Figure 1 shows that the difference between candidate groups in History was around this threshold, while for ESL and Maths 2 the standardised score differences were around double this threshold.

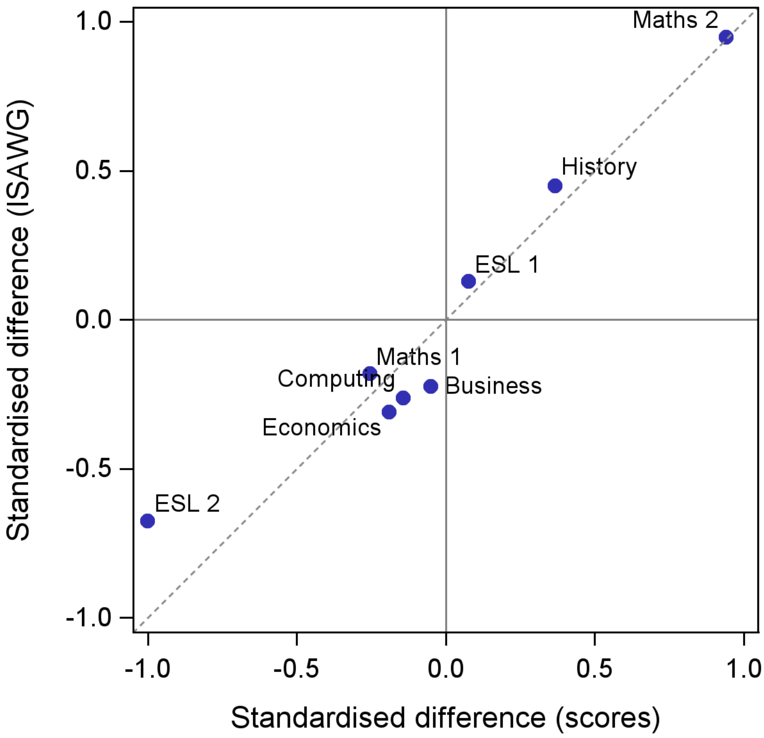

Figure 1: Comparison of standardised component differences.

3 The standardised score difference was calculated by subtracting the mean Test Y score from the mean Test X score and dividing by the pooled standard deviation of Test X and Test Y scores.

4 The standardised ISAWG difference was calculated by subtracting the mean ISAWG of Test Y candidates from the mean ISAWG of Test X candidates and dividing by the pooled standard deviation of Test X and Test Y candidates' ISAWG measures, using the ISAWG values separately calibrated for that specific Test X Test Y component pair.

Findings

How accurate was equating?

Figure 2 shows the equating outcomes for each cover sheet pair and equating method. For each score on Test X of the pair (shown on the x axis, as a percentage of the test's maximum mark), the graphs show the difference between the estimated equivalent Test Y score and actual equivalent Test Y score (equal to the Test X score itself). This provides a visual summary of how closely each estimated equating function resembled the correct (identity) function. It allows the accuracy of the different methods to be compared through the score range: in the Computing pair, for instance, the graph shows that the Rasch equating method over-estimated the equivalent Test Y scores much more than other equating methods for Test X scores between 5 per cent and 25 per cent. For higher Test X scores, on the other hand, the Rasch equated scores had similar accuracy to those estimated from the other methods.

The patterns of equating error shown in Figure 2 (that is, the deviations from the correct equating function – the identity function – as plotted on the y axis) varied by pair. For Business, equating errors were small and highly consistent between the different methods. Equating errors were also consistently small in Economics, although Figure 2 shows some separation between the two single co-component methods (which produced very similar results), and the ISAWG and Rasch methods. Equating errors were still consistently within a small range for Maths 2, although here there was more variation between the equating methods. The correlations between Business, Economics and Maths 2 scores and their anchor measures (both co-component scores and ISAWG) were all high, which would tend to support equating accuracy. The large difference in Test X and Test Y candidate abilities for Maths 2 was an apparent challenge to overcome, but this pair was nevertheless equated very accurately.

For Computing and Maths 1, pairs which had high score-anchor correlations and fairly small group differences, equating errors were small except for deviations in specific methods towards the lower end of the score range. In Computing, the larger errors occurred with the Rasch equating method, and in Maths 1, the larger errors occurred in both the Rasch and ISAWG methods.

Equating errors were slightly larger for the ESL 1, History, and particularly ESL 2 component pairs, consistent with the fact that these three pairs showed the lowest score-anchor correlations (Table 2). In the case of History and ESL 2 there were also fairly large ability differences between the Test X and Test Y candidate groups, further increasing the level of equating challenge.

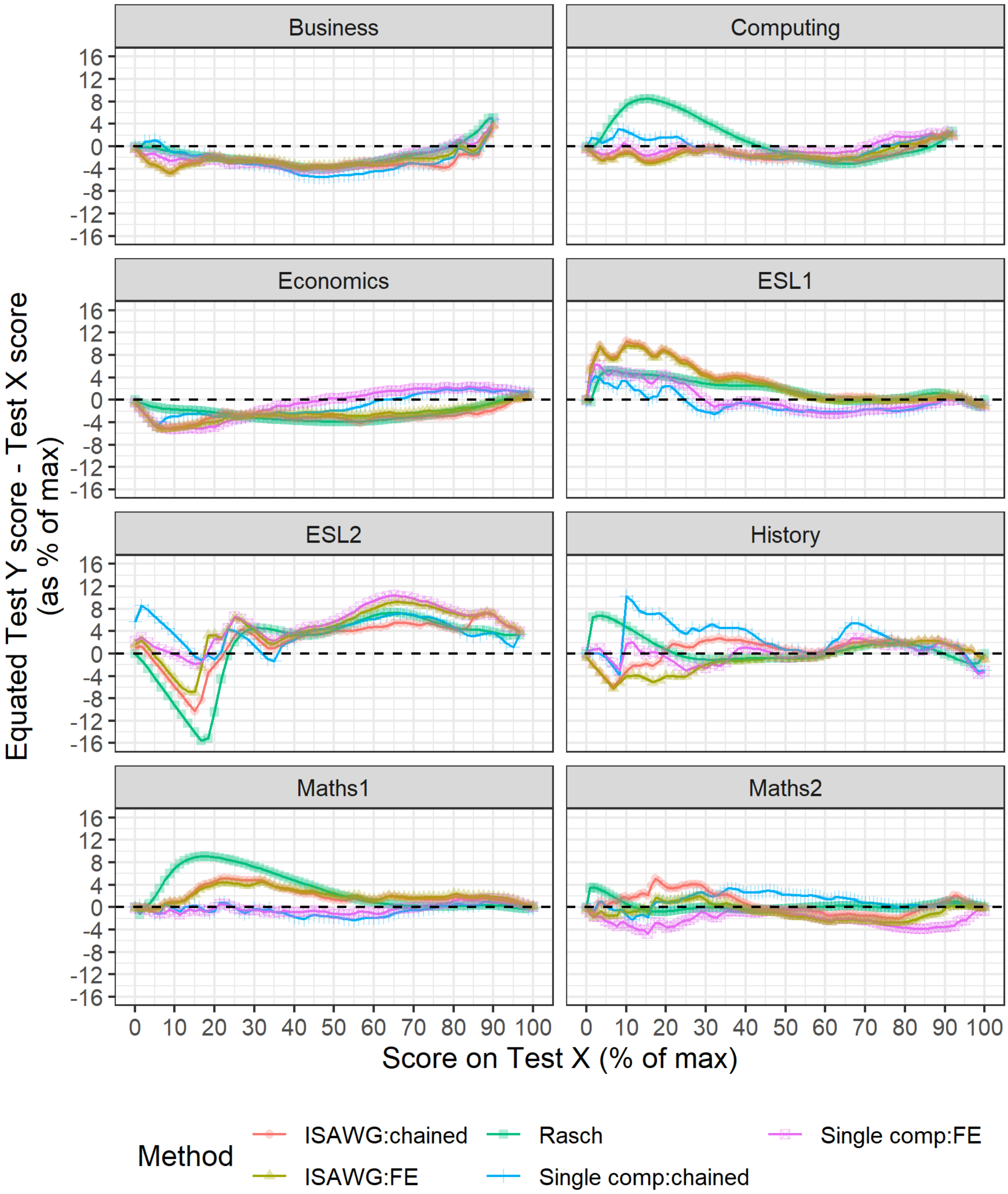

Figure 2: Equating errors, by method, against Test X score.

To summarise the size of equating errors, Table 3 reports the weighted mean absolute error of equating for each pair of equated components under each method. Weighting by the number of (Test X) candidates at each score gives priority to those parts of the score range with higher numbers of candidates. For each component pair, the lowest overall equating error (using this definition) is highlighted, which highlights that the equating method achieving the highest accuracy varied between pairs. Comparing the results for the different component pairs within each method, however, shows that each method produced its highest levels of equating error for the ESL 2 pair. Prior to running the analysis, we expected chained equating to outperform the FE method in cases where there was a large difference in group means – namely ESL 2, History and Maths 2 (see Figure 1). However, for History this was not the case, with the FE method providing more accurate results for both the ISAWG and single co-component approaches.

The final row of Table 3 shows the mean error of each method across the eight data sets. On this measure the Rasch approach was slightly more accurate than the alternatives, although the difference between them was very small.

Table 3: Weighted mean absolute errors of equating (as per cent of component max mark). For each pair, the lowest overall equating error is highlighted.

Component pair	Single comp (FE)	Single comp (chained)	ISAWG (FE)	ISAWG (chained)	Rasch
Business	2.99	4.08	2.86	3.24	2.88
Computing	1.06	1.75	1.56	1.94	2.40
ESL 1	1.56	1.68	0.43	0.61	0.53
ESL 2	7.83	5.52	6.91	4.46	5.58
Economics	1.69	1.63	2.30	2.93	2.89
History	1.15	2.49	1.17	1.48	0.95
Maths 1	0.90	0.82	2.02	1.70	1.27
Maths 2	2.72	0.86	1.51	1.12	0.36
All	2.49	2.35	2.34	2.19	2.11

Figure 3 demonstrates how the equating errors shown in Figure 2 would affect pass rates, by comparing the cumulative percentage of candidates reaching actual and equated scores. For each cover sheet pair, the Figure 3 x axis shows the cumulative percentage of Test Y candidates above a given score, for example, 40 represents the score which 40 per cent of Test Y candidates achieved or exceeded. The y axis shows the difference between this percentage and the percentage of Test Y candidates who reached or exceeded the corresponding equated score. The top left cell of Figure 3 shows that for the pair of Business assessments, the "pass rate" at a cut score achieved by 50 per cent of the Test Y cohort would have been around 10 percentage points higher using the equated cut scores from the single co-component (chained equipercentile) method shown in blue. This corresponds to the fact that the single co-component (chained) method resulted in equated scores for Business that were lower than actual scores throughout most of the score range (see top left cell of Figure 2).

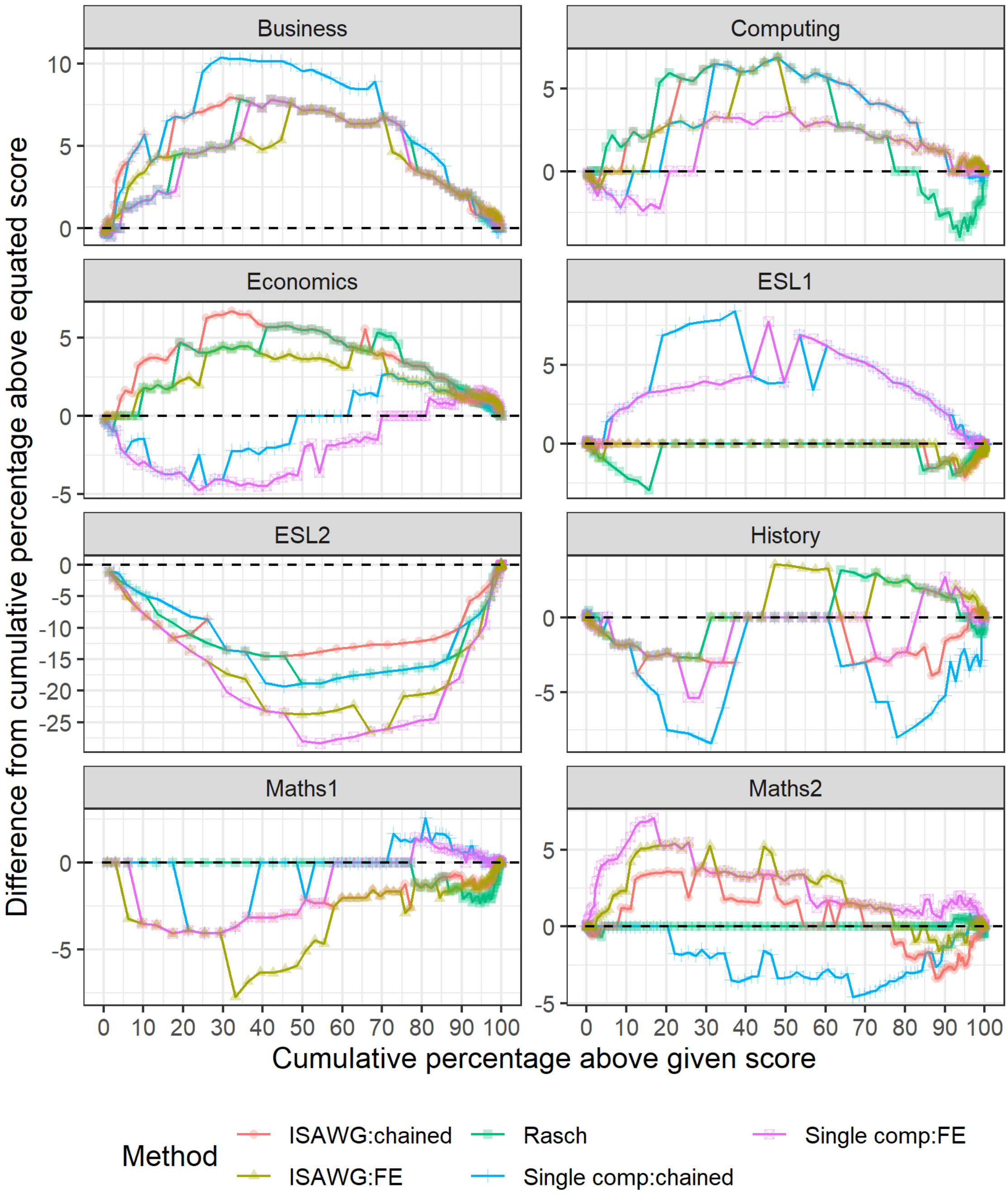

Figure 3: Differences between cumulative percentages of Test Y candidates reaching actual and equated scores.

The cumulative percentage of Test Y candidates achieving equated scores was generally within 10 percentage points of the cumulative percentage at the original score, except for the ESL 2 component pair, where differences for all equating methods exceeded this. Focusing on the cumulative percentages of candidates above a certain score highlights that differences in the "pass rates" at cut scores can become relatively large even when the absolute sizes of equating errors are modest, as seen for the Business components. Conversely, it also emphasises how large equating errors towards the extremes of the score range can have relatively little impact, since there are often few candidates with such scores. To illustrate the impact of equating errors in even more concrete terms, Table 4 reports differences between the cumulative percentages of Test Y candidates achieving the actual and equated cut scores for key grades. This shows how each method would affect the proportion of candidates reaching the key grades, if candidates were re-graded based on equated score grade boundaries. Table 4 confirms that the equated grade boundaries were exactly equal to the actual grade boundaries for a number of these grades, with consequently no difference between equated and actual pass rates (indicated in the table by "-"). These grade boundaries correspond to those parts of the cumulative distributions shown in Figure 3 where the difference between equated and actual cumulative percentages was zero. The largest differences shown in Table 4 are for the grade C boundary in ESL 2, but differences of 5–10 percentage points are seen in multiple component pairs, and for multiple equating methods.

Table 4: Differences (in percentage points) between pass rates at equated and actual grade boundaries.

Subject	Grade	Single comp (FE)	Single comp (chained)	ISAWG (FE)	ISAWG (chained)	Rasch
Business	A	5.05	10.25	5.05	7.93	5.05
	C	6.21	6.21	4.38	6.21	6.21
Computing	A	3.29	6.87	6.87	6.87	6.87
	C	1.28	1.28	1.28	1.28	-2.56
ESL 1	A	3.52	7.17	-	-	-
	C	3.92	3.92	-	-	-
ESL 2	A	-11.64	-6.68	-11.64	-11.64	-9.21
	C	-27.31	-17.55	-22.27	-13.08	-17.55
Economics	A	-4.49	-2.44	4.05	5.92	4.05
	C	-	2.22	2.22	3.26	3.26
History	A	-	-	3.52	-	-
	C	-	-2.99	-	-2.99	2.66
Maths 1	A	-	-	-2.54	-2.54	-
	C	0.77	0.77	-1.42	-0.75	-1.42
Maths 2	C	3.24	-3.47	3.24	1.63	-
	E	1.17	-3.81	1.17	-	-

Why was the performance of statistical equating so poor in one instance?

As can be seen from the previous sections, the worst equating performance was for ESL 2. As such, it is worth illustrating exactly why statistical equating has not worked in this instance. For simplicity, we will focus upon equating using a single co-component.

ESL 2 was a writing composition task that was taken as part of qualifications assessing English as a Second Language. Note that the group that took the Test X version were all located in one country whereas the group that took Test Y were mainly in another (with a minority scattered across several others). The selected single co-component was a reading comprehension test (also in common across the two assessments). Table 5 shows the performance on the components being equated and this main co-component. As can be seen, while the Test Y group were 0.6 of a standard deviation ahead in reading, they were more than one entire standard deviation ahead in writing. This mismatch between the two skills resulted in the equating error shown in Table 3 and consequent impact on pass rates shown in Table 4.

This same issue is shown visually in Figure 4. The figure shows the relationship between anchor test scores and scores on the tests being equated in each group. In order to allow the patterns to be seen more easily, the figure is restricted to a random sample of 200 candidates in each group (rather than overloading the chart with almost 30 000 points). Figure 4 shows that for the same performance on the selected anchor test, candidates in the Test Y group tended to perform much better in writing. A chart like this could lead to the misleading impression that Test Y is easier than Test X when in fact the two tests are identical.

This example serves to illustrate how, in the absence of an anchor test that actually measures the same construct as that being equated, no single statistical method can be guaranteed to perform well. The relative performances of two groups of students in a particular subject (e.g. Reading) may not reflect their relative abilities in another (e.g. Writing). As such, any method based upon co-components may occasionally give a misleading picture of the differences between groups.

Finally, it is worth noting that, although we have focused upon the use of a single co-component, neither Rasch analysis nor the ISAWG performed notably better. This demonstrates that accurate equating cannot be achieved simply by making use of more of the same kind of data, nor in altering the way in which analysis is done. At least in this instance, accuracy could only be improved if we had a better external link (ideally measuring writing ability) between the two groups of students.

Table 5: Descriptive statistics relating to equating ESL 2 component pair.

Components being equated	Total marks available	Test X group		Test Y group		Difference (in overall SDs)
		Mean	Standard deviation	Mean	Standard deviation	
Anchor test (Reading)	50	26.4	7.0	30.8	7.4	0.61
Tests being equated (Writing)	60	35.4	6.4	43.1	7.5	1.11

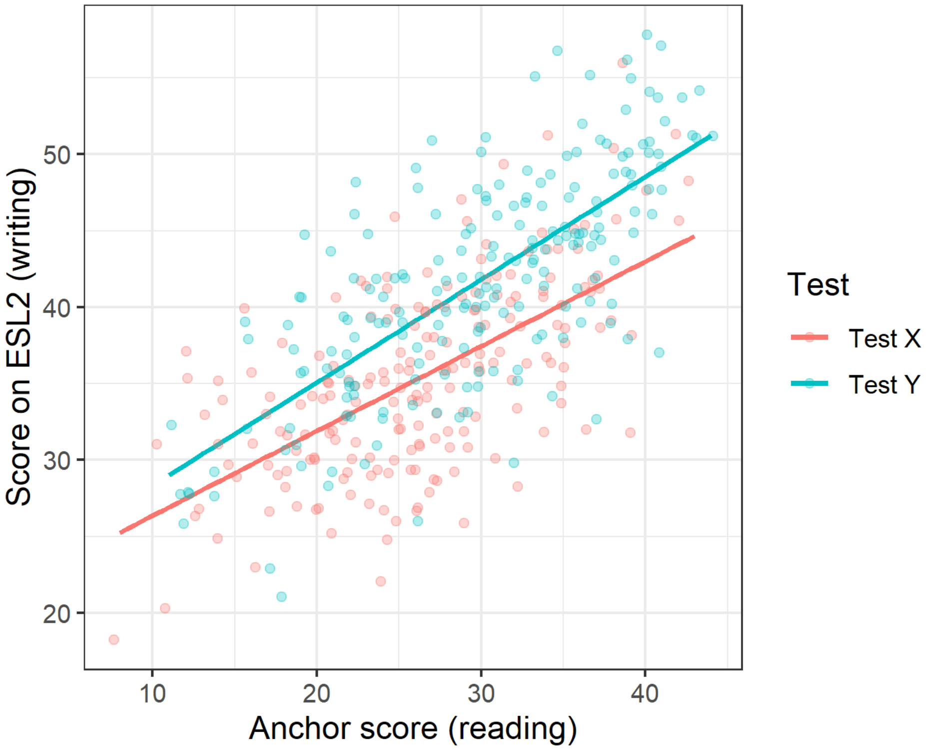

Figure 4: The relationship between performance on the anchor test (reading) and on the ESL 2 tests (writing) being equated for a sample of 200 students from each group. Regression lines are included for each group.

Conclusions

This article reported the results of equating various pairs of identical assessments. While some pairs were equated with very high accuracy by particular methods, the results showed that equating errors with real-world impact (e.g., an increase of 5–10 per cent in the proportion of students achieving a grade A) occurred even where equating conditions were apparently favourable: candidate groups were large, group differences were not extreme, and a very substantial amount of information on candidate performance in co-components was available. No single method consistently produced more accurate results than the others: the most accurate equating method varied by pair, and in fact all methods performed well for at least one component pair.

The results give further evidence that ISAWG and co-component equating methods can offer useful information towards maintaining standards. However, they also emphasise that multiple sources of information should still be considered, to make final boundary decisions.

More broadly, the results are a reminder that if applied uncritically, equating methods can lead to incorrect conclusions about the relative difficulty of assessments. In this equating exercise, Test X was not just written to the same specifications as Test Y, but was in fact identical to Test Y. However, equating between non-equivalent groups using operational data with non-random missingness as an anchor is difficult, even when we have extensive amounts of relevant information on candidates' abilities in other assessments. In the context of this study, the estimated equating relationships between pairs of identical assessments could have produced the paradoxical conclusions that assessments were both "easier" and "harder" in comparison with themselves.

References

Andersson, B., Branberg, K., & Wibery, M. (2013). Performing the Kernel Method of Test Equating with the Package kequate. *Journal of Statistical Software, 55*(6).

Benton, T. (2017). *Pooling the totality of our data resources to maintain standards in the face of changing cohorts.* 18th annual AEA-Europe conference, Prague, Czech Republic.

Bramley, T. (2011). Subject difficulty – the analogy with question difficulty. *Research Matters: A Cambridge Assessment Publication, Special Issue 2: Comparability*, 27–33.

Bramley, T., & Vidal Rodeiro, C. L. (2014). *Using statistical equating for standard maintaining in GCSEs and A levels*. Cambridge Assessment Research Report. https://www.cambridgeassessment.org.uk/Images/182461-using-statistical-equating-for-standard-maintaining-in-gcses-and-a-levels.pdf

Coe, R. (2008). Comparability of GCSE examinations in different subjects: an application of the Rasch model. *Oxford Review of Education, 34*(5), 609–636. https://doi.org/10.1080/03054980801970312

Kolen, M. J., & Brennan, R. L. (2014). *Test Equating, Scaling, and Linking: Methods and Practices* (3rd ed.). Springer.

Maris, G., Bechger, T., Koops, J., & Partchev, I. (2021). *Dexter: Data Management and Analysis of Tests*. https://CRAN.R-project.org/package=dexter

Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects*. ETS Policy Information Center.

Newton, P. (2010). Thinking About Linking. *Measurement: Interdisciplinary Research & Perspective, 8*(1), 38–56. https://doi.org/10.1080/15366361003749068

Newton, P. E. (2012). Making sense of decades of debate on inter-subject comparability in England. *Assessment in Education: Principles, Policy & Practice, 19*(2), 251–273. https://doi.org/10.1080/0969594x.2011.563357

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology, 15*(2), 201–293.

Wiberg, M., & Branberg, K. (2015, Jul). Kernel Equating Under the Non-Equivalent Groups With Covariates Design. *Applied Psychological Measurement, 39*(5), 349–361. https://doi.org/10.1177/0146621614567939

Progress in the first year at school

Chris Jellis (Cambridge CEM)

Introduction

Children normally start school at the age of four in state-maintained schools in England. This year is known as the Reception Year. In the time up to their starting school children will have had a wide range of learning experiences, and a good teacher will want to find out as quickly as possible what new pupils know and can do. It is also important to the teacher and other teachers within the school to know what progress they make in the first year at school. This paper draws upon an analysis of data produced by a computer-based assessment (BASE – not an acronym) that teachers carried out with their pupils shortly after the children entered full-time education in the Reception class, and that was repeated at the end of the year. The paper describes what children could typically do when they started school and the progress children typically made in their first year at school.

Background to the BASE assessment

In 2015, the Department for Education (DfE) introduced the idea of a statutory baseline assessment for use in the Reception Year in state-funded schools. In response to this requirement, the Centre for Evaluation and Monitoring (CEM) created a new baseline assessment named BASE which was accepted as one of three assessments that schools could use to fulfil their statutory duties. The assessment was offered to all CEM schools currently using an earlier baseline assessment and to new schools that chose to use BASE from the three options available. The first year of assessment was the academic year beginning in September 2015.

The DfE's decision to introduce baseline assessment was not without its opponents. Critics of early assessment (Bradbury, 2019) frequently argue that testing children results in them being "labelled". The BASE assessment was never designed to "label" children, but was constructed to allow teachers to discover the skills and knowledge already possessed by children and to help them to build upon these.

Following the pilot year, a small number of items in the BASE assessment were reviewed and replaced in response to user feedback, making the first stable year for BASE the academic year starting in September 2016. Since its inception, the BASE assessment has also been offered to other CEM assessment users, such as those in independent schools, international schools, and schools in Scotland. As the number of state-maintained schools using the assessment has diminished, the rise in other schools taking up the assessment has changed the demographic representation of the sampled population. Despite the changing composition of schools and pupils taking the BASE assessment, it has remained popular, being taken by an average of 26 000 pupils each year since 2016.

The BASE assessment

BASE is administered within the first few weeks of the child starting in the Reception class. It is taken on a computer on a 1:1 basis with a teacher or other suitable adult. A cartoon character on the computer screen asks questions and the child answers, either verbally or by pointing to an object on the screen. The response is then marked on screen by the teacher.

The assessment is not fully computer adaptive but uses a simple "three wrong and move on" algorithm, ensuring that if questions get too hard, further questions of greater difficulty are not asked. Once a child's level of ability is reached in a particular section the assessment moves on to the next topic. Reports are then generated showing which questions the child answered correctly and a score showing where they stand in the overall ability range for the national BASE cohort starting in that year. The child is then assessed again at the end of the year and measures of progress can be established.

The BASE assessment consists of over 200 questions in 13 sections. The difficulty of these questions ranges from questions appropriate for typical 3-year-olds up to questions appropriate for typical 6-year-olds. Due to the adaptive nature of the assessment, only the most able children will see all the questions in each section. When the children are assessed again in the end-of-year assessment (EOY), typically children will not see questions they have already answered correctly in the start-of-year assessment (SOY) but will be moved on to questions they have not yet seen. The questions chosen for analysis in this investigation were drawn from the initial section in each assessment area (maths, literacy etc.,), ensuring that most of the children would be offered these questions.

Table 1: The sections of the BASE assessment analysed.

Section name	Number of questions
Concepts about Print	10
Repeating Words	9
Vocabulary	14
Letter Recognition	26
Word Recognition	6
Shapes	5
Counting and Numerosity	6
Numbers	23
Numeracy 1	9
Total	**108**

What do 4-year-olds typically learn in school?

When children start school, it may be the first time that they have been in an academic environment. Some, though, may have attended a nursery, playgroup or kindergarten where there was some formal teaching occurring. For others, the child's parents or siblings may have involved themselves in the child's learning. Every Early Years programme is designed to teach children the basic skills they need to make sense of the world around them and to access formal education as they grow up. To this end, Early Years education concentrates on early number and literacy work. The sections of the BASE assessment are grounded in educational research to provide teachers with important information about the children they teach.

Mathematics

The development of a sense of number is the foundation of all mathematics. Psychologists have found that children are born with a basic concept of numerosity, and that very young children will show surprise or concern when one toy is surreptitiously removed from a small number of toys they have been looking at (Feigenson et al., 2002; Langer et al., 2003). It follows then, that learning mathematical concepts does not start at school and some children have been introduced to single digits and even to numbers with two or three digits, so there are BASE questions to cover that area. Two things are being addressed here; firstly, that the child can distinguish between the single digits by their shape, but secondly, that they know a name for that digit. This part of the assessment does not assess the concept of number itself, but digit recognition can form the groundwork for understanding place value and how number systems work.

Counting is also an area where children can develop an early sense of number. Counting combines digit identification with the concept of cardinality, that is, the number of items in a set. Children begin to understand that counting involves visiting each element in a set and assigning a number to it. The final number they reach is also the size of the set (Schaeffer et al., 1974; Nunes & Bryant, 2009). This is by no means simple, and as the size of the set increases, children develop strategies to keep track of the elements they have counted and those they have yet to count. Once they have established strategies for counting groups of objects, children normally move on to problems involving 'counting on' or counting back', that is, they are beginning to understand the concepts of addition (counting on) and subtraction (counting back) (Nunes & Bryant, 2009). These skills lead on directly to sharing (division) and counting groups (multiplication).

The ability to recognise shapes is an important precursor to the understanding of geometry and there are subtleties to a child's learning in this area. Young children can often distinguish between a square and a triangle but understanding that a square rotated through 45° is still a square, can often be too hard for them (Tall, 2013). At first, children are distinguishing shapes by their gross morphology, but as they learn more about shapes, they start to understand the nomenclature based on the number of sides (hexagon, pentagon etc.).

Reading

Learning to read is a complex process which is initiated by the child developing an understanding that print conveys information. By reading to young children and observing where their attention lies it is possible to capture some fundamental behaviours relating to visual perception, mental processing and motor development. Asking children to point to parts of the story uses a combination of these basic behaviours to make sense of the text being shared. This is the basis of Concepts About Print developed by Marie Clay in New Zealand (Clay, 1989). Clay established some fundamental skills that young readers (and pre-readers) develop. Among these are the correspondence between each word they read and the word on the page, directionality (in Western texts, pages are read from left to right) and the relationship between letters, words and sentences. These ideas are shared, along with the recognising and naming of individual letters and ultimately words, in the concept of 'Emergent Literacy' advocated by Whitehurst and Lonigan (1998) but based on Marie Clay's doctoral dissertation entitled 'Emergent Reading Behaviour' (Clay, 1966). Emergent literacy is seen as a continuum which starts with pre-reading. The concept also considers the interdependence of reading, writing and oral language in the development of literacy.

The ability to recognise individual letters and associate them with specific sounds is fundamental to learning to read. Most schools use a specific phonics scheme to teach children the sounds of individual letters and how they are modified when they appear together. A pilot of the synthetic phonics approach in schools in Clackmannanshire, Scotland, (Johnston & Watson, 2005) showed a promising result and, following a review by Sir Jim Rose ("The Rose Report", Rose, 2006), a former director of inspection at Ofsted, English state schools were heavily encouraged to adopt a synthetic phonics scheme. However, a systematic review carried out by Carol Torgerson and her colleagues at the Universities of York and Sheffield (Torgerson et al., 2006) found no statistically significant evidence for the use of synthetic phonics. Opponents of the phonics method of teaching point out that phonics works very well when there is a one-to-one relationship between letters and sounds in languages such as Italian, Greek and Spanish (Goswami, 2008), but that languages like English, where the letter "a" takes different sounds in common words such as "car", "talk", "cat" and "make" do not lend themselves as well to a phonological approach. The BASE assessment avoids these debates by recognising that children starting school are likely to have been taught some letters and letter sounds, but not necessarily using a specific phonics scheme. As such, it accepts the name or the sound as a correct answer to the recognition of a letter.

Word recognition provides an insight into a child's letter recognition, phonics, and their development of reading. Simple consonant-vowel-consonant (CVC) words such as "cat' or "dog' can be read by using a phonological approach, but given time, these words may be recognised without sounding them out. As Nation and Snowling (2004) point out, there is a distinction between decoding and word recognition, and reading fluency depends on automatic word recognition of familiar words.

Building a wide vocabulary is also extremely important in the development of reading skills. The vocabulary section of the BASE assessment concerns itself with the ability of the child to make sense of the world around them by naming the things they see. A number of studies (Lee, 2011; Hayiou-Thomas et al., 2012; Bleses et al., 2016), have established a link between vocabulary size and future achievement.

The responses to BASE questions in the areas described above can provide evidence of the progress any child has made. This evidence can be in the form of the final score (how many questions they answered correctly), but also in a qualitative way (how familiar was the child with letter or numbers, adding or reading simple words?). This evidence provides a richer picture of the child's skills and understanding of these basic concepts, which is of great benefit to teachers planning their lessons.

Method

BASE item-level data for the academic years beginning 2016 (32 047 individual pupils), 2017 (22 127 individual pupils) and 2018 (16 457 individual pupils), (total 70 631 individual pupils) was obtained for both SOY and EOY assessments. Although data was available from many different schools, this analysis was restricted to results obtained from state-maintained schools in England only. Initial sections of the assessment were chosen for analysis, covering the first stages of Literacy and Numeracy. These are shown in Table 2.

Table 2: Initial sections of the BASE assessment.

Area	Concept	How it is assessed
Literacy	Concepts about Print (CAP)	The child is asked to point to individual letters, words, where to start reading and some punctuation in a page of text shown on the screen.
Literacy	Letter Recognition	The child is shown letters of the alphabet, some as lower case and some as upper case. An acceptable response is either the sound or the name of the letter.
Literacy	Word Recognition	Here the child is shown very short (two or three-letter) words and asked to read them out loud.
Literacy	Vocabulary	The child is shown a series of pictures and asked to point to specific objects within each picture.
Numeracy	Shapes	The child is shown a picture containing many different shapes and is asked to point out specific ones (square, triangle etc.).
Numeracy	Number Recognition	The child is shown single-digit, then two-digit and higher numbers and asked to name them.
Numeracy	Counting	The child is asked to count items of varying numbers starting from four and increasing to numbers in the thirties.
Numeracy	Numeracy 1	The child is asked to do simple arithmetic such as addition or subtraction.

The data from all three years was combined and the items calibrated with an IRT (Rasch) model. For ease of interpretation, in the results section the item difficulties are presented as estimates (based on the model) of the percentage of the entire population of test-takers that would have answered them correctly if all the items had been presented to everyone. A higher percentage value therefore represents an easier item.[1]

Results

Concepts about Print

Figure 1 shows the item difficulties for the start and end of the year for all children combined when they were asked to point to specific items on a page of text.

1 A Marginal Maximum Likelihood (MML) algorithm was used to estimate the IRT model parameters in order to account correctly for the non-random nature of the missing data arising from the partially adaptive item selection algorithm in the BASE test. For further details see Eggen & Verhelst (2011).

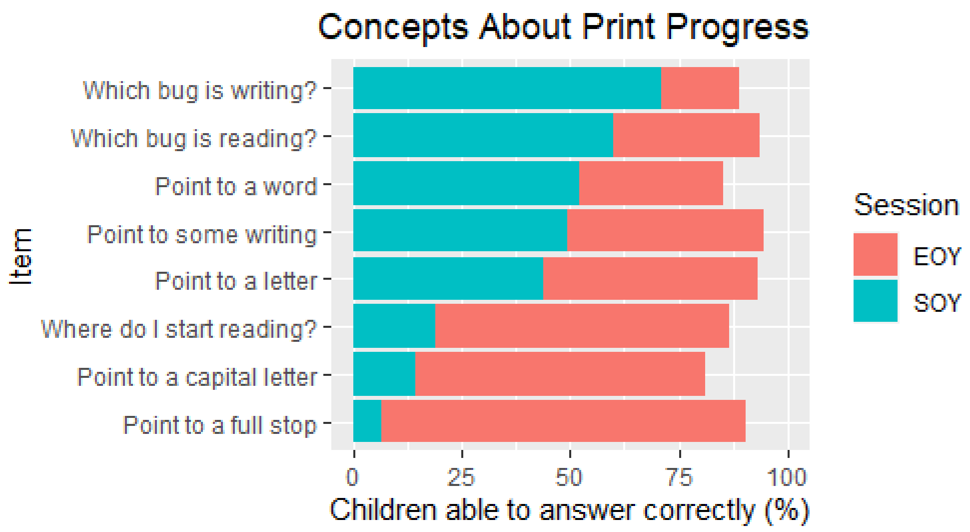

Figure 1: Concepts about Print items, difficulties and progress.

The easiest item required the child to point to someone who was writing. 71 per cent of pupils could do this at the start of the year, rising to nearly 89 per cent at the end of the year. The next easiest was pointing to someone who was reading. 59 per cent of pupils could do this at the start of the year and around 94 per cent by the end of the year. Fewer than 20 per cent of children entering school knew about full stops, capital letters or where to start reading, but by the end of the year 81 per cent or more were able to do this.

Letter Recognition

Figure 2 shows the difficulty values for the start and the end of the year for each letter of the alphabet. Note that some letters are capitals, and some are lower case. For expediency, the assessment does not ask children whether they recognise each letter as lower case and then again in upper case as it would be extremely time consuming. Moreover, the time for which young children can concentrate on a single task type is limited.

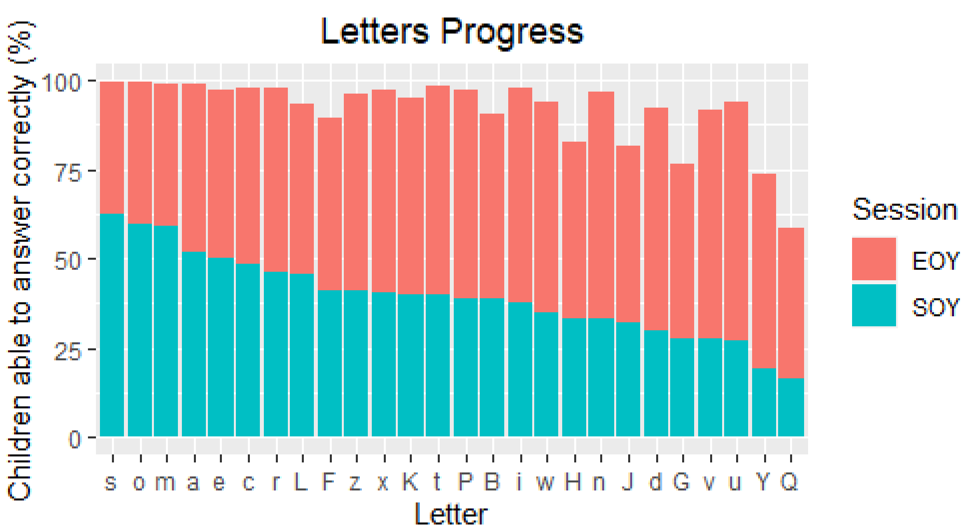

Figure 2: Letter Recognition items, difficulties and progress.

At the start of the year the most recognised letters were lower case "s", lower case "o" and lower case "m". Around 60 per cent of children were able to recognise them. The hardest letters for children to recognise were upper case "Y" and upper case "Q". Fewer than 20 per cent of children could recognise those letters. By the end of the year most children could name almost all the letters of the alphabet. What is often overlooked is that the children are not only learning the shapes of the letters and their corresponding sounds and names, but as they will be taught using the phonics approach, they will be actively combining newly learnt letters to form simple words.

Word Recognition

Figure 3 shows the words that children are asked to read. Although the words themselves seem remarkably simple, learning to read involves a great deal of mental gymnastics. The reader must know the sounds associated with each letter and can then combine them to produce an overall sound – the complete word. Learning to read normally starts with CVC words (consonant-vowel-consonant), moving on to CVCC words or words such as "see" where the "e" sound is modified when two are together.

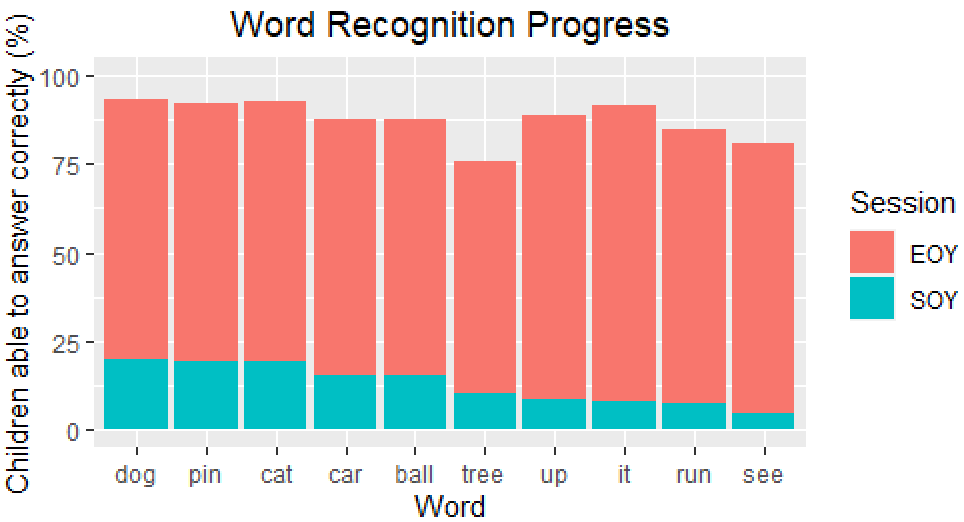

Figure 3: Word Recognition items, difficulties and progress.

Being able to read whole words was a skill that only around 20 per cent of children could do on entry to the school. The easiest item for those who could read simple words was "dog" and the hardest was "see". However, after a year in school, between 75 per cent and 93 per cent of children could read these simple words. Of these, the most difficult at EOY was "tree", containing as it does a combination of consonants "tr" and a double "e".

Vocabulary

The ability to put names to objects is fundamental to learning about the world around you. Figure 4 shows the difficulty of the vocabulary items asked in the BASE assessment and the proportion of children that could recognise that item in a picture onscreen at both the start and the end of the year.

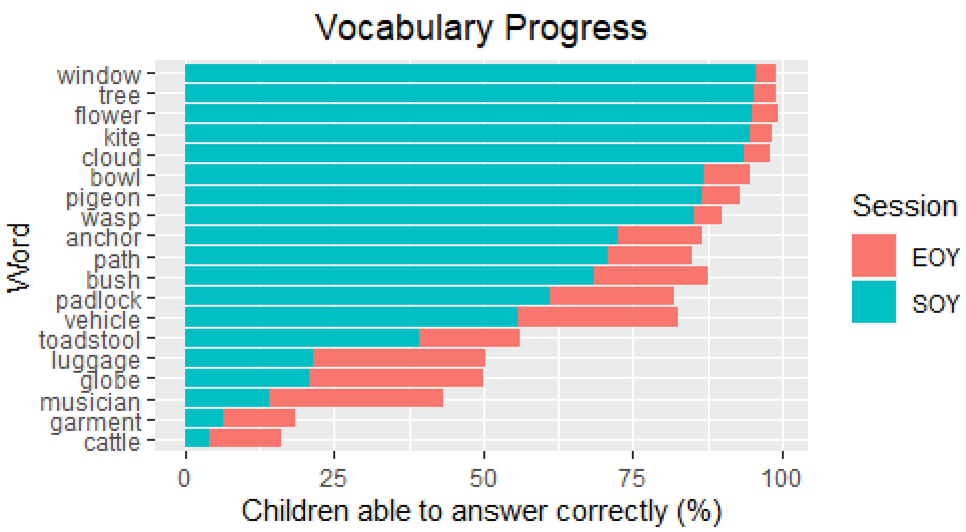

Figure 4: Vocabulary items, difficulties and progress.

English vocabulary contains many synonyms, some of which cause issues with vocabulary tests. For one child it is a "pan", for another child it is a "pot". However, at the start of the year most children (around 95 per cent) could point to a window, a tree, a flower, a kite, and a cloud. Very few children (less than 10 per cent) understood the words "garment" and "cattle". Even at the end of the year "garment" and "cattle" were extremely challenging words, but a significant proportion of the children in the group were able to answer them correctly (18 per cent and 16 per cent respectively).

Shapes

Figure 5 shows the names of the shapes that the children were asked to point to.

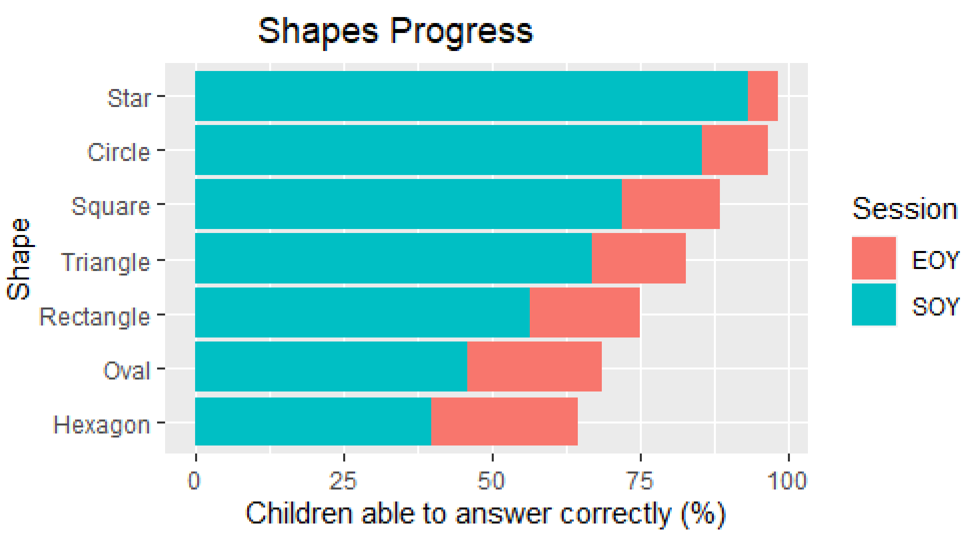

Figure 5: Shape recognition items, difficulties and progress.

At the start of the year the most recognised shapes were the star and the circle. Over 85 per cent of children could point to these. The most difficult was the hexagon, although around 40 per cent of children knew this shape. At the end of the year hexagon and oval were still the hardest shapes to name, but over 64 per cent of children could do this.

Number Recognition

The children were asked to name numbers as they appeared on the screen. As with most parts of the assessment, if they started to get the answers wrong the program would move on to another section of the test covering a different topic. The numbers selected include some tricky items, but due to the adaptive algorithm built into the assessment, fewer children would see these.

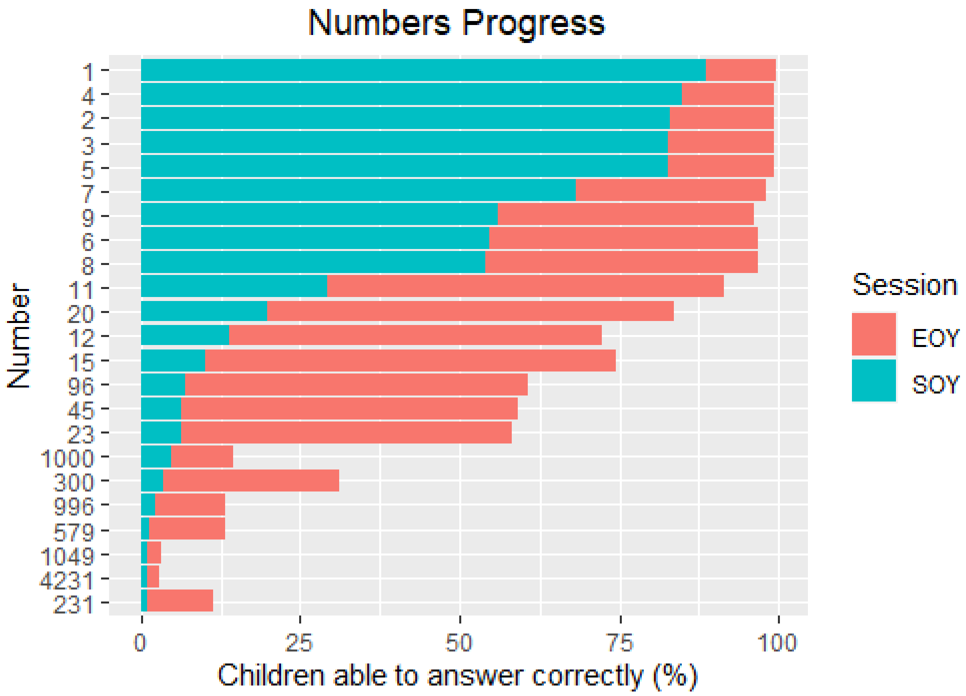

Figure 6: Number recognition items, difficulties and progress.

Interestingly, the difficulty of recognising individual digits was found not to correspond with their numerical order. At the start of the year around 80 per cent of children knew the numbers 1 to 5, but only 55 per cent of children could recognise the number 6. Slightly more children were able to recognise the numbers 7 and 9. Fewer than 10 per cent of children recognised numbers greater than 15. The end of the year showed a vast improvement and more than 58 per cent of children could name one- and two-digit numbers. As would be expected at this age, three- and four-digit numbers still proved to be challenging for these children. Some results stood out, particularly the large rise in the number of children that could name the numbers 300 and 231. This could be evidence of more children understanding the concept of place value combined with the lower value digits with which they might be more familiar.

Counting

Children were asked to count the spots on the back of a ladybird, and later, count the number of ladybirds on a page. This method is preferred over counting on a number line as children can be observed pointing to each spot as they count, and it is easier to see those who count the same spot twice, miss a number or do not know where to stop.

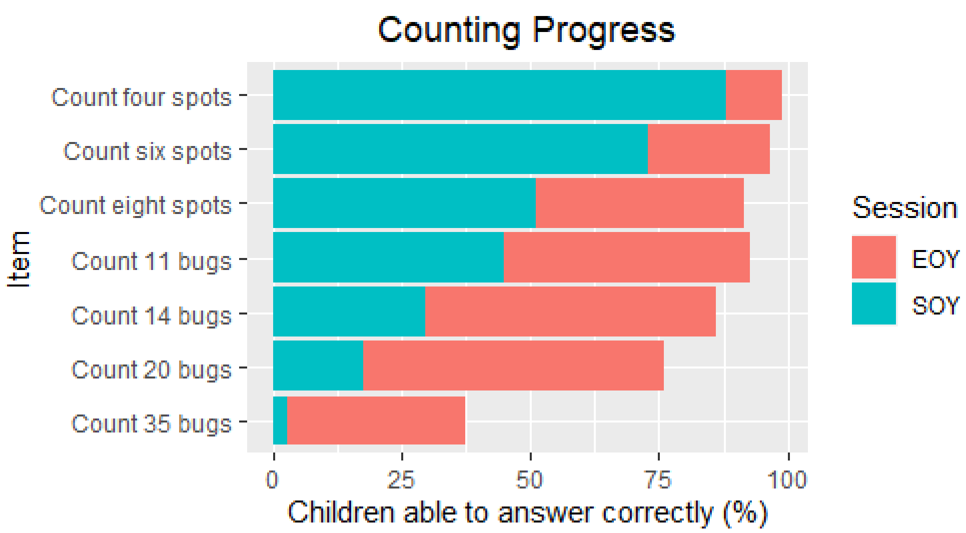

Figure 7: Counting items, difficulties and progress.

At the start of the year many children (88 per cent) were able to count four items and slightly fewer (73 per cent) could count six items. By the end of the year just over 75 per cent could count 20 bugs and 37 per cent could count 35 bugs, a large rise from just under 3 per cent at the start of the year.

Numeracy

This section of the assessment allows the teacher to observe how children performed when they attempted addition and subtraction problems.

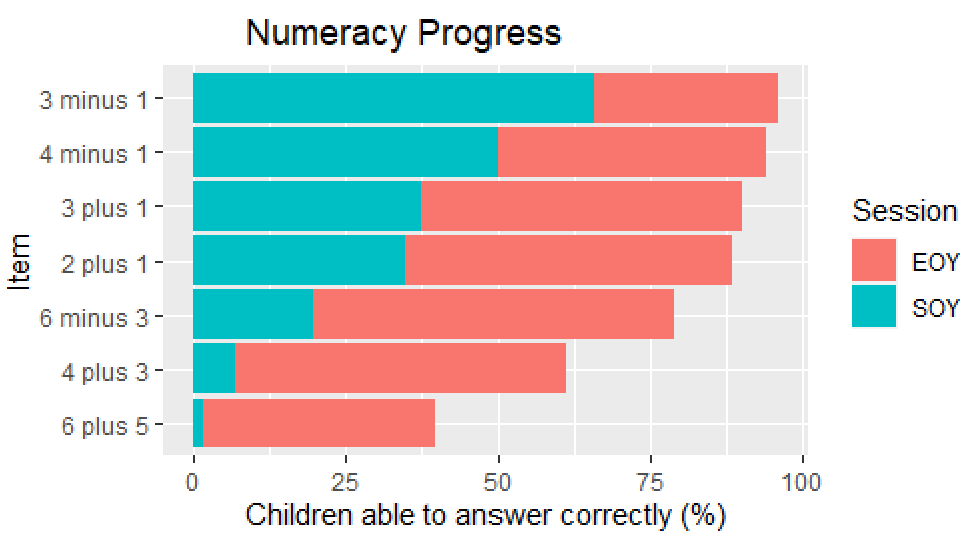

Figure 8: Basic mathematics items, difficulties and progress.

Initially, just over 50 per cent of the group were able to subtract 1 from a small number. Far fewer (around 36 per cent) could add 1, and even fewer (<10 per cent) could add a number other than 1. At the end of the year there was remarkable progress. Almost everyone could add or subtract 1, and over 60 per cent could add or subtract 3. Around 39 per cent could add 5 to a number, up from less than 2 per cent at the start of the year.

Discussion

It is clear from the results shown above that many children arrive in school with a great deal of academic knowledge and many skills. They can count, recognise shapes and simple letters, and have a growing vocabulary. The results also show the astonishing progress that four-year-olds make once they are in school. For example, at SOY fewer than 20 per cent of children could read simple CVC (consonant-vowel-consonant) words, but 85 per cent of children could read those words at EOY. Similarly, fewer than 20 per cent of children could calculate 6 minus 3 at SOY, but at EOY the number that could do so had risen to around 80 per cent.

As our results show, around 88 per cent of children starting school can count four things, but just under 3 per cent can count 35 things. At the end of the year this changes radically: over 75 per cent can count 20 items and nearly 40 per cent can count 35 items. Again, it is important to clarify that what is being assessed is not the ability just to count to an arbitrary number, but to engage with items in several patterns and layouts – a much harder problem.

Some issues do arise though. In schools, letters and letter sounds are often taught initially using synthetic phonics schemes which begin with the letters "s", "a", "t", "p", "i" and "n" (Jolly Phonics, Letterland, FFT Success for All Phonics). The rationale for this is that these are common letter sounds and many simple words can be created by combinations of those letters. As may be seen from the analysis of BASE data, although the letters "s" and "a" are reasonably well known among school starters, the letters "t", "i" and "n" are not. (BASE asks children to recognise the upper case letter "P".) Ironically, it may be those letters that children are most likely to recognise that are problematical. Phonics teaching applies very specific sounds to each letter that are unlikely to have been taught by parents. It may be that the most common letters cause more difficulties in phonics learning than those that are yet to be learnt, because children may have been taught the letter with a different "sound" than that taught in synthetic phonics lessons.

We also see a link between digit recognition and counting. More than half of the children entering the Reception class could recognise the digits 1 to 9, but Figure 6 shows that the order of difficulty does not follow the natural order of the numbers. For instance, fewer children could recognise the number 6 than could recognise the numbers 7 or 9. Following on from this, counting (the application of number) shows that most children could count four and six spots on the back of a ladybird, but counting eight or more was much more difficult.

The area where most pupils starting school struggled was in arithmetic. Although around half the pupils could take 1 away from a small number, far fewer could carry out additions of any type, or subtraction of numbers larger than 1. This type of insight is extremely useful in informing pedagogy. Why is it that children who have understood the concept of subtraction find it more difficult to subtract numbers larger than 1? Some studies (Carey, 2001; Rips et al., 2008) have suggested that this ability is limited by short-term memory and attention. If they are counting back in single digits from the given number, then is there too much cognitive overload to keep track of the number of times they must count back, and the result?

Conclusions

Children often learn many basic skills before they enter school, and the sources and extent of this learning can be very varied. Some will be taught by their parents or other relatives and carers. Some will attend playgroups or nursery schools. Even those without these advantages may be able to learn a great deal from the excellent learning resources readily available from a range of digital providers.

This diversity of experience prior to entering formal schooling is why it is important to establish a baseline for children entering the Reception class. Equally valuable is repeating the assessment at the end of the first year in school. It can provide a measure of relative progress for each child, and of the whole group.

The BASE assessment provides an opportunity to assess a class of children objectively and comprehensively in a range of basic skills. The results are often surprising, as initial perceptions of what children can or cannot do are frequently challenged. It is easy to overlook this, and treat every child the same, rather than aiming to differentiate groups according to their individual learning needs. Similarly, it is vital to realise how important early learning opportunities are to growing minds.

As the data shows, young children can make remarkable progress when placed in an environment that encourages active and engaged learning. Indeed, the effects can be extremely far reaching. A longitudinal study by Peter Tymms and his colleagues (Tymms et al., 2018) found that “Membership of an effective Reception class/school was associated with a boost in attainment that was still apparent at age 16.”

Acknowledgements

Thanks to Tom Benton and Matthew Carroll for determining the most appropriate analysis method for the BASE data.

References

Bleses, D., Makransky, G., Dale, P., Højen, A., & Ari, B. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics, 37*(6), 1461–1476. https://doi.org/10.1017/S0142716416000060

Bradbury, A. (2019). Datafied at four: The role of data in the 'schoolification' of early childhood education in England. *Learning, Media and Technology, 44*(1), 7–21.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenisis. *Mind & Language, 16*(1), 37–55.

Clay, M. M. (1966). *Emergent reading behaviour.* Unpublished doctoral dissertation, University of Auckland, New Zealand.

Clay, M. M. (1989). Concepts about print in English and other languages. *The Reading Teacher, 42*(4), 268–276.

Department for Education. (2021). *Statutory framework for the Early Years Foundation Stage: Setting the standards for learning, development and care for children from birth to five.* Department for Education.

Eggen, T. J., & Verhelst, N. D. (2011). Item calibration in incomplete testing designs. *Psicológica, 32*(1), 107–132. https://recyt.fecyt.es/index.php/psi/article/view/11343

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological science, 13*(2), 150–156.

FFT Success for All Phonics, https://fft.org.uk/phonics/

Goswami, U. (2008). Learning to read across languages: the role of phonics and synthetic phonics. In K. Goouch & A. Lambirth, A. (Eds.), *Understanding phonics and the teaching of reading: critical perspectives,* (pp. 124–143). McGraw-Hill Education.

Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2012). The etiology of variation in language skills changes with development: A longitudinal twin study of language from 2 to 12 years. *Developmental Science, 15*(2), 233–249.

Johnston, R., & Watson, J. (2005). *The effects of synthetic phonics teaching of reading and spelling attainment: A seven year longitudinal study.* https://www.webarchive.org.uk/wayback/archive/3000/https://www.gov.scot/Resource/Doc/36496/0023582.pdf

Jolly Phonics, https://www.jollylearning.co.uk/

Langer, J., Gillette, P., & Arriaga, R. I., (2003). Toddlers' cognition of adding and subtracting objects in action and in perception. *Cognitive Development, 18*(2), 233–246.

Letterland, https://www.letterland.com/
Nation, K., & Snowling, M. J., (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342–356.

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32*(1), 69-92.

Nunes, T., & Bryant, P. (2009). *Key understandings in mathematics learning. Paper 2: Understanding whole numbers.* Nuffield Foundation. https://www.nuffieldfoundation.org/wp-content/uploads/2019/12/P2.pdf

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rips, L. J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and Brain Sciences, 31*(6), 623–642.

Rose, J. (2006). *Independent review of the teaching of early reading.*

Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology, 6*(3), 357–379.

Tall, D. (2013). The Foundations of Mathematical Thinking. In *How Humans Learn to Think Mathematically: Exploring the Three Worlds of Mathematics* (Learning in Doing: Social, Cognitive and Computational Perspectives, pp. 33–49). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139565202.007

Torgerson, C., Brooks, G., & Hall, J. (2006*). A systematic review of the research literature on the use of phonics in the teaching of reading and spelling.* DfES Publications.

Tymms, P., Merrell, C., & Bailey, K. (2018). The long-term impact of effective teaching. *School Effectiveness and School Improvement, 29*(2), 242–261. https://doi.org/10.1080/09243453.2017.1404478

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child development, 69*(3), 848-872. https://doi.org/10.1111/j.1467-8624.1998.tb06247.x

What are "recovery curricula" and what do they include? A literature review

Martin Johnson (Research Division)

Introduction

Curriculum review, the systematic study of curriculum-related documents, allows us insight into the social context of education. Such study tells us something about the conditions, aspirations and objectives that are important when a curriculum was developed. This point is conveyed by Stabback (2016, p. 6) when he states: "the curriculum … embodies a society's educational aims and purposes". Similarly, changes to educational curricula are also indicative of concurrent changes in the surrounding social context (Swiss National Science Foundation, 2017). Observations suggest that such changes may reflect responses to unfolding situations such as economic crisis (Ragnarsdóttir & Jóhannesson, 2014), ideological shifts (Dichter, 2012; Hallama, 2020) or calls for decolonisation (Lidher, McIntosh & Alexander, 2020; Winter, 2018).

Recovery curricula are developed in response to educational disruption and have an important role in educational rebuilding. "Recovery" has many associations, including medical, economic, and nation building (following conflict), although the common component of all emergencies is that they require an educational response to be developed in situations that are fluid and often unforeseen. This literature review draws from documents that cover all these forms of emergency, and includes academic papers, government policy and guidance documents, non-governmental organisation (NGO), charity and United Nations (UN) agency reports, and educationalists' blogs.

In the first two decades of the 21st century there appears to have been a heightened interest in recovery curricula[1], and so a study of the character of these curricula can also tell us something about the educational conditions over this period.

1 The Scopus database contains 19 documents with the term "Recovery Curriculum" in the title. 18 of these documents were published between 2000 and 2022.

What do we mean by "educational curricula"?

Before looking at recovery curricula in detail it is useful to consider what is meant by educational curricula. Curriculum is a contested and often misunderstood concept. Simple conceptualisations that imply a course of study are insufficient for understanding the complex processes of schooling. Some educationalists favour a definition that considers curriculum to be an umbrella term denoting the totality of the learning experience of children and young people in school (Priestley, 2019). This mirrors the characterisation expressed by John Kerr, who defined the curriculum as "all the learning which is planned and guided by the school, whether it is carried on in groups or individually, inside or outside the school" (Kerr, 1968, p. 16). Both of these conceptualisations ensure that the concept of curriculum includes the "what", "how" and "why" of learning (learning objectives; content; the way that learning is structured; strategies for instruction; and assessment). This all-encompassing definition can be seen in some contemporary national initiatives. For example, in Scotland's *Curriculum for Excellence* the curriculum is "the totality of all that is planned for children and young people throughout their education" (Scottish Government, 2008, p. 13), and in the new curriculum for Wales "A school's curriculum is everything a learner experiences … It is not simply what we teach, but how we teach and crucially, why we teach it" (Welsh Government, 2020b).

According to Porter and Smithson (2001) it is important to distinguish between "intended" and "enacted" curricula. Intended curricula include overt, documented, stated curricular ambitions, and these are likely to be found in published policy texts or guidance documents. In contrast, enacted curricula include the lived or received learning experiences that can be evidenced from observational data. For methodological reasons I use a relatively narrow conceptualisation of the curriculum, one with a greater emphasis on curricular aspirations than on observed curricular experiences. An important part of my literature review covers policy and guidance documents, and these tend to convey the intended rather than the enacted curriculum (see Creese, Gonzalez & Isaacs, 2016, for a study with similar aims and concerns). Moreover, the contemporaneous character of the review literature means that it is unlikely that there would be evidence of curriculum impact (since it can sometimes take years for the full effects of a curriculum initiative to achieve its impact). Despite this, my review methodology did allow me to gather information about a variety of curriculum contexts, which it would be difficult to achieve through other approaches.

Before looking in detail at the review methods, I will discuss the concept of educational recovery.

What do we mean by "recovery" in education?

Definitions of educational recovery and the role of curricula in that process are relatively opaque in the academic literature[2]. Dictionary definitions of recovery focus on ideas around returning to a previous or "normal" state or regaining possession or control of something lost. In educational terms, such loss might relate to a diminished access to learning, or reductions in expected levels of attainment as a consequence of some man-made or natural disruption to the education system. As a consequence, educational recovery appears to link with ideas around reinstating access to established curriculum objectives and content, and these will differ according to specific cultural and historical contexts.

Mentions of recovery curricula in relation to educational disruption are found across a variety of contexts. These contexts include post-conflict situations (Barakat et al., 2013), school closure (Carpenter & Carpenter 2020), and natural disaster (Akbar & Sims, 2008). These curricula also span educational phases, from early years (Goddard, 2020) to secondary level (Sherwood, 2020).

Interestingly, references to recovery curricula appear across a range of national contexts, including England (Brennan, 2020; Dickens, 2020), Scotland (McLaughlin, 2020), Ghana (GhanaWeb, 2020) and the United States (Jawor, 2020). This international dimension is understandable. Recovery has a global dimension as these issues tend to cross national boundaries and can often lead to human displacement. For example, the contemporary scale of emergency human displacement is considerable, with children accounting for around half of the estimated 26 million refugees reported in 2019 (UNHCR, 2019b).

Recovery in education: from a specialised to a universal concern

These demographics help to explain why the field of Education in Emergencies has emerged and grown since the turn of the century. Reflecting on some recent intergovernmental responses to emerging crises, it appears that educational recovery is a central concept. For example, in outlining their mission UNESCO states "(a)s the UN lead agency for Education, UNESCO plays an active role in promoting education as a part of emergency response and for long-term recovery" (UNESCO, 2017). Looking at some specific initiatives across other UN agencies, we can see similar messages. In response to the effects of Hurricane Matthew in Cuba in 2017 UNICEF explains that "Many children … needed early psychosocial recovery and new learning materials. UNICEF supported the Ministry of Education's recovery efforts by donating cases with school kits, primary education kits and recreation kits, all of which have helped children continue to learn" (López Fesser, 2017). In their review of their refugee education initiatives, UNHCR outlines that "For refugees, [education] is … the surest road to recovering a sense of purpose and dignity after the trauma of displacement. It is – or should

2 An initial search of academic literature reveals many references to "recovery education". This form of recovery tends to have a specifically medical focus, such as educational programmes dealing with mental health or alcohol dependency issues (e.g., see, Moos & Moos, 2006; Reid et al., 2020). This article does not deal with this medicalised concept of recovery but focuses on curricula that are designed in response to a general disruption in educational provision.

be – the route to labour markets and economic self-sufficiency, spelling an end to months or sometimes years of depending on others" (UNHCR, 2019b, p. 5).

This shared, intergovernmental interest led to the development of the Inter-agency Network for Education in Emergencies (INEE) in 2000 "as a communication mechanism for advocacy, knowledge sharing and the distribution of materials to promote improved collaboration and effectiveness in the context of education in emergencies" (Mendizabal & Hearn, 2011, p. 109). The centrality of educational recovery (and the influence of curricula on this) is clear in the INEE Minimum Standards for Education which they claim are "A global tool that articulates the minimum level of educational quality and access in emergencies through to recovery" (INEE, 2010). The Standards cover guidance on learning access, curricula and pedagogy, and policy formation (among other things).

There are clear parallels between the educational recovery work of the INEE and its partners and the educational responses to the COVID-19 pandemic. The emergency status of education in the pandemic is highlighted by data from the UNESCO Institute for Statistics. This data shows that there have been 130 country-wide school closures during the pandemic, affecting around 990 million learners (UNESCO, 2020).

The disruption to education systems as a response to the pandemic has led to an interest in the concept of a recovery curriculum and around the nature of the sorts of curricular responses that may be appropriate to this situation. In the next section of this article, I outline how I gathered and analysed information about recovery curricula to get a picture of this particular type of curriculum.

Review method

Fink (2010) and Heyvaert, Hannes and Onghena (2016) outline six curriculum review stages. These stages include research question formulation, database identification, search term definition, literature selection, literature reviewing, and synthesis of the research literature data. In my review I wanted to identify (1) the objectives and content that are included in recovery curriculum documents and (2) any evidence for the efficacy of such curricula. I included seven document sources[3], which then led to a snowball approach (e.g., see Atkinson & Flint 2001) that picked up some additional secondary sources. All the documents were published in English. I used three sets of search terms ("Recovery + Curriculum"; "Catch up + Curriculum", and "Education + Emergency + Curriculum"). I also limited the searches to research from the year 2000 as this coincides with the establishment of the INEE. This search identified 38 documents, and these included academic papers, government policy and guidance documents, NGO, charity and UN agency reports, and educationalists' blogs.

I used MAXQDA (VERBI Software, 2021) to collate and code these documents. These sources were tagged to identify the country, the educational phase, the

3 Clarivate Web of Science™; University of Cambridge iDiscover; Taylor & Francis Online; Wiley Online Library; ARD *Curriculum Watch* Data; Education Sub Saharan Africa Research Database; Inter-agency Network for Education in Emergencies website.

scope (i.e., national or regional), and the form of emergency that they related to. The 38 documents covered five different (although sometimes overlapping) emergency types (Figure 1). Health Emergency was the most commonly covered emergency type. Civil Conflict and Migration shared some overlaps (since one is often a spur for the other) but I kept these categories separate since their link is not a necessary one. I also had a "General" category of documents as some sources covered a variety of emergencies.

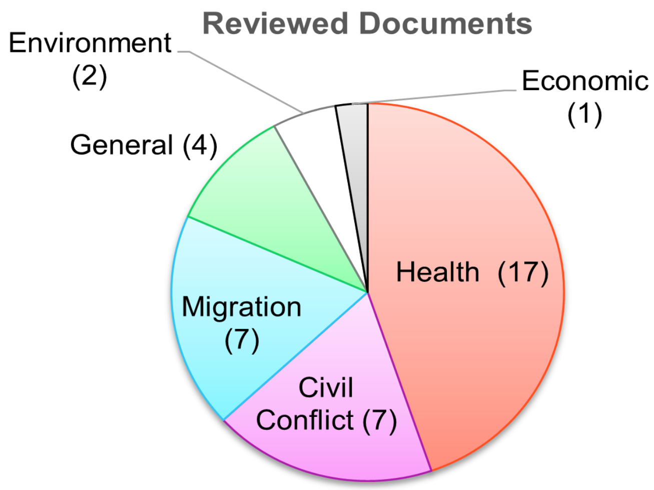

Figure 1: Emergency types covered in the source documents.

There were some apparent relationships between the curriculum document publication dates and the type of emergency that they were designed to deal with (Figure 2). The Health Emergency documents all emerged in 2020 (linked with the onset of the COVID-19 pandemic), the Civil Conflict documents were published between 2000 and 2010 (focusing on education in East Timor-Leste, former Yugoslavia, the Democratic Republic of Congo, Liberia and Rwanda), while the Migration documents were published between 2009 and 2019 (focusing on South Africa, Kenya, Lebanon, Libya, Syria and Turkey).

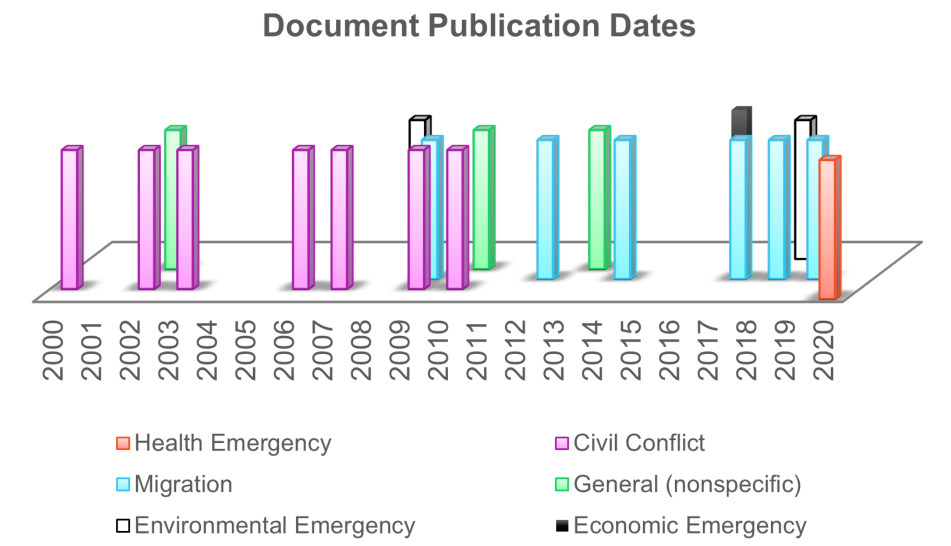

Figure 2: Source document publication dates.

The source documents were then reviewed, and a set of thematic codes developed and applied to each document. This coding then formed the basis of my synthesised analysis.

Analysis

My coding analysis suggested that the documents included information that fell into five different thematic areas. These areas were Curriculum Objectives; Pedagogy; Curriculum Content; information about the Curriculum Development Process, and Efficacy. My coding also allowed me to identify the most common information that was related to each thematic area (Table 1).

Table 1: Curriculum review themes.

Objective	Pedagogy	Content	Development	Efficacy
• Support wellbeing • Support teacher readiness • Support learner readiness	• Parent/ community involvement • Contextual-isation • Cross-curricular	• Language/ literacy • Maths/ numeracy • Health and wellbeing	• Rich resource development • Prioritisation	• Participation rates • Educational outcomes • Behaviour • Integration

Curriculum Objectives

It was very common for the curriculum objectives in recovery curricula source documents to relate to supporting learner wellbeing and teacher and learner readiness. Coding analysis showed that these elements were not discrete from each other but were linked holistically.

Supporting learner readiness had a social and emotional component, as well as connections with specific learning content (such as access to core foundational learning concepts in Language/Literacy and Maths/Numeracy). Supporting learner readiness commonly linked with building learner resilience and preparing learners to deal with uncertainty and new situations. Supporting learner readiness also linked with the idea of taking steps to reduce learner anxiety and reinforcing wellbeing through helping learners to build relationships, e.g., "a recovery approach ... enables students of all ages to reconnect and rebuild emotional resilience with a strong focus on relationships" (Gray, 2020).

There was also a common link between learner readiness and with ensuring that teaching focused on the fundamentals of core skills and knowledge which would support the learner for later learning. These core skills and knowledge are termed "priority outcomes" by the The Inter-agency Accelerated Learning Working Group (AEWG) (2020, p. 2). This linkage between readiness and core learning areas was reflected in advice from the UK Department of Education around the COVID-19 lockdown in England: "it may be in the best interests of a year 11 pupil to discontinue an examined subject because the school judges that, for example, they would achieve significantly better in their remaining subjects as a result, especially in GCSE English and mathematics" (Department for Education, 2020).

Pedagogy

The most common pedagogic messages in the source documents indicated that there was a dual focus on (1) the need to flexibly adapt education to local conditions, and (2) the need to consider learners' social development through the way that the curriculum was delivered. The Pedagogy codes that appeared in most recovery curricula source documents related to involving parents and the local community. For example, Almasri et al., (2019, p. 95) highlight how the accepted "basic principles of education in emergencies involve adopting a community-based approach".

There were also common references to the need for flexible approaches to the delivery of the curriculum (including decisions involving school management, timetabling, governance etc.), or to contextualising the curriculum to learner needs. This point has clear links with the social and emotional components of the recovery curriculum objectives covered in the last section. One curriculum document expressed this as "(a)llow time for individual children and families to tell their lockdown stories. You can adapt curriculum content to reflect this or be flexible with delivery" (Cornerstones Education, 2020).

Recovery curricula were generally cross-disciplinary in structure. By encouraging educators to establish links between different knowledge areas, the documents were drawing educators' attention to the possibility of organising and delivering the curriculum in flexible ways. This issue also linked closely to the explicit goal of focusing on core knowledge content. The literature highlighted how the key elements of language and mathematical core knowledge can be integrated across multiple areas of learning, for example, "Understanding informational texts and identifying important information helps learners in science and social studies, as well as language arts. Creating graphs and interpreting data helps learners in science and social studies, as well as mathematics. Analysing the meaning of a question or problem and responding to it are skills that can be applied to any subject area" (The Inter-agency Accelerated Learning Working Group (AEWG) 2020, p. 5), and "learners should have opportunities to develop and apply these [literacy, numeracy and digital competence] skills across the curriculum" (Welsh Government, 2020, p. 7).

The concern for learners' social development was also to the fore in the source curriculum documents. References to parent and community involvement were found most often in primary education phase documents, and these also reinforced the central importance of play, relationship building, and home links for the education of younger learners. In Northern Ireland, the Department of Education captured this in their COVID-19 advice to schools, "(i)nitially, in primary and special schools in particular, it is likely that activities will often focus on getting pupils used to routines and safe behaviours, interacting with others within the rules and building the ability to engage with activities and sustain concentration. Play and social interaction within the protective bubble of the class are centrally important for younger children" (Northern Ireland Department of Education 2020a, p. 6).

Curriculum Content

The curriculum content that appeared in most recovery curricula source documents related to Language & Literacy, Maths & Numeracy, and Health & Wellbeing.

Language & Literacy

Language & Literacy was particularly important in recovery curricula for several reasons. It was commonly associated with catch-up objectives which focused on bringing learners up to speed with expected levels of attainment. During the COVID-19 pandemic emergency, English schools, and particularly those in the primary sector, focused on aspects of potential language loss (e.g., "All the primary school leaders told us that they are concentrating hard on reading, including phonics. Many leaders explained that they wanted to make sure that if there have been any losses in learning, particularly in reading, these are quickly put right" (Ofsted, 2020, p. 3)). Language & Literacy was also linked with supporting social inclusion and peacebuilding initiatives in recovery curricula. It was noted that language learning is a component of recognising diversity and overcoming ethnic tensions in some post-conflict contexts (Obura, 2003, p. 88), and in creating a new shared national identity (Shah, 2009, p. 5).

When looking at the literature from international contexts it is important to consider whether Language & Literacy refers to home language (the learner's first language) or the host country language in which learning is taking place (which could be an additional language for the learner). To pull this issue apart I separated out the contexts in which the Language & Literacy references were made (i.e., "Health Emergency", "Migration", "Civil Conflict", etc.). My analysis showed that the distinction between home and host language mainly occurred in the Migration Emergency sources. The lack of discourse around the language of learning for the Health Emergency curricula suggested that language choice was not an issue for education systems where migration was not a factor.

In displacement contexts, such as in Migration Emergency and some Civil Conflict situations, language learning policy had a different emphasis and was marked by insecurity. The recovery curriculum literature suggests that decisions about the language of instruction were influenced by whether the curriculum objective was to support learner repatriation to the home system or to integrate them into the host system, and these decisions are not always clear cut. The United Nations High Commissioner for Refugees (UNHCR, 2012) favours host language instruction as this supports learner integration into the society where they are located. Despite this ambition, some have argued that this policy can shut down learners' opportunities, "choosing one [i.e., the home language rather than the host language] might effectively foreclose opportunities in the other", since teaching in the home language might "better prepare refugees to repatriate but might come at the expense of education in exile" (Karam et al., 2017, p. 460). There are also some concerns about the ability of displaced teachers to deliver education through a host language (Karam et al., 2017, p. 456), and that the use of the host language can erode learners' cultural identity (Karam et al., 2017, p. 457) as language “carries notions of identity, culture, power and control” (Pausigere, 2009, p. 59).

Advocates for home language instruction argue that this better supports their repatriation once a crisis is over. Teaching through the home language can also benefit younger learners” access to core knowledge (e.g., “Using a child’s first language or mother tongue for initial literacy instruction in school enhances pupils’ achievement” (UNICEF, 2000, p. 6); “Pupil achievement is enhanced if pupils first become literate in their mother tongue” (UNICEF, 2000, p. 8). It is also recognised that learning through another language can be a significant hindrance to attending education (Sinclair, 2002 p. 10). Despite this ambition, some argue that home language policy risks maintaining the migrants “on the social and economic periphery of and in inferior positions within the host state” (Pausigere, 2009, p. 12).

Maths & Numeracy

Language & Literacy references in curriculum sources often sat alongside references to Maths & Numeracy, suggesting that these areas formed core curriculum components. It was notable that different education programme types highlighted the central importance of literacy and numeracy. These programme types included non-formal education programmes (Karam et al., 2017; Kagawa, 2005), refugee education programmes (Halstead & Affouneh, 2006; Pausigere, 2009; Smith, 2013), accelerated education programmes (The Inter-agency Accelerated Learning Working Group (AEWG) 2020), and COVID-19 recovery guidance (e.g., Scottish Government/Riaghaltas na h-Alba, 2020; Welsh Government, 2020; Northern Ireland Department of Education, 2020a, 2020b).

This reinforces the point that Language & Literacy and Maths & Numeracy were widely considered to be the common principal components of core knowledge across a variety of contexts, ensuring that they were the focus of many recovery curricula. This is articulated by the Inter-agency Accelerated Learning Working Group: “A condensed curriculum does not teach all subject areas faster. Rather, it centres teaching and learning activities on ‘priority outcomes’. Priority outcomes describe essential skills and knowledge that are transferrable across multiple subject areas: reading, writing, mathematics, critical thinking, and problem solving. Priority outcomes give learners the tools they need for future, self-directed learning” (The Inter-agency Accelerated Learning Working Group (AEWG), 2020, p. 2).

It was also noteworthy that these components of core knowledge were mainly linked to Primary education documents, suggesting that they are key elements that need to be covered in the earliest phases of a recovery curriculum. UNICEF conveyed this in their guidance on curriculum design, “Curriculum must specify adequate instruction time for basic subjects, especially language development and mathematics in primary grades” (UNICEF, 2000, p. 7).

Health & Wellbeing

Health & Wellbeing was another content area that appeared more than most others in recovery curricula. It was generally associated with content to do with with peace, conflict resolution and citizenship education. This association reflected a perspective that education was important in helping learners to recover from the trauma related to conflict, with the curriculum "…supporting the development of refugee education programmes that meet the psychosocial needs of children and adolescents and promote health, safety, environmental awareness, and skills of conflict-resolution and citizenship" (Sinclair, 2002, p. 90). Recovery curricula content also reflected the health issues that were prevalent in some conflict contexts, where it was important to "Deploy literacy, numeracy, life skills, and other emergency education curricula, including on health, hygiene promotion, HIV prevention, environmental education, peace education, and other appropriate emergency themes" (Smith, 2013, p. 48).

"Non-core" content elements

It is also important to note that these foundational skills were not the only areas of learning included in the recovery curricula. For example, the documents from the UK were explicit in their appeal for recovery curricula to "teach an ambitious and broad curriculum in all subjects" (Department of Education, 2020); and for "learners [to] have learning experiences that span a broad curriculum and include opportunities to develop a breadth of understanding and a range of knowledge and skills that then lead to further depth" (Welsh Government, 2020a, p. 8). In addition to the core subject areas discussed above, the following areas of learning and development were also represented within the reviewed documents:

Non-core elements	
Creative Arts	Nature/Outdoor/Environmental Education
Digital Competences	Peace, Conflict Resolution, Citizenship
History/Humanities	Physical Development
Human Rights Education	Religious Education
Learning Skills/Metacognition	Science
Life Skills [Problem Solving, Creativity, Critical Thinking]	Social & Emotional Development

Curriculum Development Process

The reviewed literature also included some information and guidance on how to construct a recovery curriculum. This information related to (1) the importance of prioritisation, and (2) the role of resource development. Prioritisation involved decision making around identifying the elements of curriculum content that were the most important in a particular context. For example, OECD guidance for education planners responding to COVID-19 identified the need to "Re-prioritize curriculum goals … Define what should be learned during the period of social distancing" (Reimers & Schleicher, 2020, p. 5). The development of resources to back up the recovery curriculum (to support teacher readiness) was also mentioned in many of the source documents (e.g., Reimers & Schleicher, 2020; UNICEF, 2000; Northern Ireland Department of Education, 2020a; Department for Education, 2020).

Looking at the ways that curriculum development connected with other issues in recovery curricula, resource development was most commonly associated with supporting teacher readiness and supporting social inclusion. While the links between resources and teacher readiness are alluded to above (e.g., Reimers & Schleicher, 2020), there was also recognition in the sources that development programmes that aimed to shift traditional and perhaps less inclusive curricula required adequate support materials if teachers were to transform established practices (Sinclair, 2002; Obura, 2003).

When considering reprioritisation, it was most common for this to focus on core knowledge, and Language & Literacy in particular. For example, observations of recent changes in Primary teacher practice in England highlighted that "they were teaching most of the subjects they usually teach, though many have reordered topics within subjects. Primary schools were giving even more attention to reading than usual" (Ofsted, 2020, p. 2).

The reviewed documents also suggested that when engaging in flexible curriculum delivery (e.g., reordering curriculum coverage to support cross-disciplinary teaching) educators needed to ensure that they maintained a transparent and sequential content structure. This transparency was helpful for supporting learner catch up in core knowledge (UNICEF, 2000) and teacher readiness (Shah, 2009).

Efficacy

I analysed the documents to find indications of positive outcomes from different recovery curricula. It is noteworthy that there was relatively little in the reviewed documents that evidenced where any particular curriculum had resulted in tangible benefits. This coheres with other observations "that there is an absence of robust evidence-based research for all educational interventions in crisis-affected zones" (Almasri et al., 2019, p. 96).

Measures of efficacy varied across the recovery curriculum contexts. For post-conflict and migration contexts efficacy indicators focused on increasing learner participation rates (Barakat et al., 2013; Shah, 2009), raising educational outcomes (Shah, 2009; UNICEF, 2000), and improving learner integration (Awada et al., 2018). For environmental emergency contexts efficacy focused on encouraging positive learner behaviours (Liberty, 2018).

Looking at how indications of efficacy linked with other elements of the recovery curricula, educational outcomes were most frequently linked with curriculum components that supported teacher readiness, implicating the provision of good quality support resources. Guidance from UNICEF highlights that "For pupils to achieve, teaching must be effective. This means that education systems must support teachers in developing appropriate teaching strategies for helping all children to achieve" (UNICEF, 2000, p. 6).

Where positive integration was considered to have occurred in some post-conflict migration emergencies, there was an association with curriculum objectives that set out to support social inclusion and/or curriculum content that dealt with Human Rights Education. For example, one claim about the reforms in Lebanon was that "the 1997 curriculum was reformed upon the end of the civil war, and it helped unite the Lebanese again to a certain extent after incorporating Human Rights Education" (Awada et al., 2018, p. 44).

Discussion and conclusions

When considering the messages from my review, it is important to recognise some important methodological limitations. Earlier, I outlined how educationalists conceptualise curriculum in a broad sense, with curricula not being simply captured as documented intentions in texts but also existing in an enacted form, and this has implications for curriculum study. The curriculum concept that I use is a narrow one. For this review I had to take a more limited conceptualisation since my sources for review contained a number of policy and guidance documents, and a limitation of policy study is that it tends to focus on intentions rather than actual practices (which can differ dramatically).

At the same time, my review methodology did allow me to gather information about a variety of curriculum contexts, which would be difficult to achieve through other approaches. My review demonstrates that there is a lot of recent interest in the concept of recovery curricula, and this raises the question as to whether it is "a thing" or "many things". By gathering a collective pool of documents, I was able to see that, when taken together as a whole, there are some common features that pertain to recovery curricula. Many of the documents shared a focus on similar objectives (supporting learner wellbeing, learner readiness and teacher readiness) and prioritised an emphasis on covering core, foundational learning content in areas such as Language & Literacy and Maths & Numeracy. There was also a coherence across the documents in terms of how resources were expected to play a part in teacher preparation, particularly where recovery entailed teachers changing their already established practices.

There were also some differences between the recovery curricula, reflecting the different contexts for which they were designed. Health Emergency documents had a greater concern with guidance than those for other emergency types (which dealt more with design issues). This characteristic might reflect the fact that the Health Emergency documents dealt exclusively with the COVID-19 emergency and mainly addressed already well-developed education systems. Curricula for other emergency types might also be considering a variety of issues, such as the complete (re)design of education systems as they cope with displaced learner (and teacher) populations.

This point is reflected in the way that the objectives for the Health Emergency curricula focused on encouraging educational continuity (e.g., supporting learners' re-engagement with a previous curriculum, helping learners to catch up on missed learning from that curriculum, and helping planners to refocus on the key components of that curriculum). This contrasts with the narrative for other curricula which may require severe restructuring as they may contain the roots of conflict. These differences also feed through to the curriculum objectives for the recovery curriculum documents across the emergency types, where we can see a contrasting emphasis on knowledge coverage in Health Emergency documents compared with social inclusion and teacher readiness in Civil Conflict/ Migration Emergencies.

On reflection, it appeared that divergences across the documents were found at the level of aims and objectives, pressing home the point that the context of the emergency that the recovery is designed to deal with is the overriding feature that influences the shape of a recovery curriculum. This means that although recovery is a common concept, the nature of what constitutes recovery depends on the nature of the emergency that instigates it.

Finally, I looked for evidence of the efficacy of recovery curricula and found relatively little in the reviewed documents to support claims that any particular curriculum had resulted in tangible benefits. It is important to recognise that this should not be taken to mean that there were no learning gains or other benefits related to recovery curricula, just that the evidence to evaluate or quantify any such gains was not found. This observation coheres with others who have noted an absence of robust evidence-based research for interventions in emergency contexts (e.g., Almasri et al., 2019). It is tempting to consider why this might be the case. It is possible that the often complex and fast-moving conditions in which emergency education initiatives are developed and enacted make it difficult to capture evidence of progress, with most effort being devoted to the delivery of education rather than its evaluation. This focus on managing education delivery during the course of an emergency may be more about adjusting education to new realities rather than about returning education to its to pre-emergency trajectory.

The lack of focus on evaluation may also be understandable as studying curriculum impact is highly complex, even in non-emergency situations. I have already alluded to how curriculum can be interpreted very broadly, covering learning across a variety of locations. For example, schooling incorporates both formal, timetabled learning activities as well as out-of-hours, extracurricular activities that influence learning outcomes. This reinforces the point that informal aspects of schooling should also be considered to be aspects of the curriculum (Kelly, 2004, p. 7). A natural extension of this argument is that the study of curriculum should consider activities beyond schooling. Studies suggest that out-of-school activities can impact learning in some cultural historical contexts (for example see the work of Ólafsson (2013) on home-based literacy learning expectations in 19th century Scandinavia, or the work of Pozzetta and Mormino (1998) and Tinajero (2010) on the “el lector” literacy learning traditions in Cuban cigar factories).

While reiterating the importance of building evaluation processes into the recovery curriculum design phase, the need to take a broad perspective of curriculum into account makes this challenging. Such an evaluation would need to consider a broad array of evidence that links to the objectives of the curricula.

References

Akbar, R., & Sims, M. J. (2008). Surviving Katrina and Keeping Our Eyes on the Prize: The Strength of Legacy and Tradition in New Orleans's HBCU Teacher Preparation Programs. *Urban Education*, *43*(4), 445–462. https://doi.org/10.1177/0042085908315943

Almasri, N., Tahat, L., Skaf, S., & Masri, A. A. (2019). A digital platform for supervised self-directed learning in emergencies: The case of the Syrian crisis. *Technology, Pedagogy and Education*, *28*(1), 91–113. https://doi.org/10.1080/1475939X.2019.1568293

Atkinson, R., & Flint, J. (2001). Accessing Hidden and Hard-to-Reach Populations: Snowball Research Strategies. *Social Research UPDATE*, *33*, 1–4.

Awada, G., Diab, H., & Faour, K. (2018). A call for curriculum reform to combat refugees crisis: The case of Lebanon. *The Curriculum Journal*, *29*(1), 43–59. https://doi.org/10.1080/09585176.2017.1400450

Barakat, S., Connolly, D., Hardman, F., & Sundaram, V. (2013). The role of basic education in post-conflict recovery. *Comparative Education*, *49*(2), 124–142. https://doi.org/10.1080/03050068.2012.686259

Brennan, J. (2020, November 20). Schools must plan to stay ahead of the Covid curve. *Times Educational Supplement*. https://www.tes.com/magazine/article/schools-must-plan-stay-ahead-covid-curve

Carpenter, B., & Carpenter, M. (2020). A Recovery Curriculum – Evidence for Learning. *Evidence for Learning*. https://www.evidenceforlearning.net/recoverycurriculum/

Committee on the Rights of the Child. (2008). *Report of the Committee on the Rights of the Child on its General Discussion on the Right of the Child to Education in Emergencies Situation*. Office of the United Nations High Commissioner for Human Rights (OHCHR). https://www.right-to-education.org/resource/report-committee-rights-child-its-general-discussion-right-child-education-emergencies

Cornerstones Education. (2020, July 1). What is a recovery curriculum and how can primary schools implement it? *Cornerstones Education*. https://cornerstoneseducation.co.uk/news/what-is-a-recovery-curriculum-and-how-can-primary-schools-implement-it/

Creese, B., Gonzalez, A., & Isaacs, T. (2016). Comparing international curriculum systems: The international instructional systems study. *The Curriculum Journal*, *27*(1), 5–23. https://doi.org/10.1080/09585176.2015.1128346

Department for Education. (2020). *Guidance for full opening: Schools*. Department for Education. https://www.gov.uk/government/publications/actions-for-schools-during-the-coronavirus-outbreak/guidance-for-full-opening-schools

Dichter, H. L. (2012). Rebuilding physical education in the Western occupation zones of Germany, 1945–1949. *History of Education*, *41*(6), 787–806. https://doi.org/10.1080/0046760X.2012.745171

Dickens, J. (2020, May 11). 6 schools findings from the new COVID-19 recovery strategy. *Schools Week*. https://schoolsweek.co.uk/speed-read-6-schools-findings-from-the-governments-covid-19-recovery-strategy/

Fink, A. (2010). *Conducting Research Literature Reviews: From the Internet to Paper*. SAGE.

GhanaWeb. (2020, September 28). *Prof. Anamuah-Mensah advocates for 'recovery curriculum programme' for primary school pupils*. GhanaWeb. https://www.ghanaweb.com/GhanaHomePage/NewsArchive/Prof-Anamuah-Mensah-advocates-for-recovery-curriculum-programme-for-primary-school-pupils-1071154

Goddard, C. (2020, September 1). *EYFS Best Practice in Schools—On the bright side*. Nursery World. https://www.nurseryworld.co.uk/features/article/eyfs-best-practice-in-schools-on-the-bright-side

Gray, S. (2020, July 30). Guest Blog: Reconnection to Recovery and Resilience – how a relationship-focused approach can help schools to prepare for full re-opening. *The Thrive Approach*. https://www.thriveapproach.com/news/relationship-focused-approach

Hallama, P. (2020). Struggling for a Socialist Fatherhood: "Re-educating" Men in East Germany, 1960–1989. *East European Politics and Societies*, *34*(4), 817–836. https://doi.org/10.1177/0888325419891200

Halstead, J. M., & Affouneh, S. J. (2006). Educating the human spirit in times of conflict: The case of emergency education in Palestine. *International Journal of Children's Spirituality*, *11*(2), 199–215. https://doi.org/10.1080/13644360600797172

Heyvaert, M., Hannes, K., & Onghena, P. (2016). *Using Mixed Methods Research Synthesis for Literature Reviews: The Mixed Methods Research Synthesis Approach*. SAGE Publications.

INEE. (2010). *Minimum Standards for Education: Preparedness, Response, Recovery*. Inter-agency Network for Education in Emergencies. https://inee.org/standards

Jawor, A. (2020, February 4). *Dickinson-Iron Intermediate School District implements new math curriculum*. Https://Www.Uppermichiganssource.Com. https://www.uppermichiganssource.com/content/news/Dickinson-Iron-Intermediate-School-District-implements-new-math-curriculum--567557661.html

Kagawa, F. (2005). Emergency education: A critical review of the field. *Comparative Education*, *41*(4), 487–503. https://doi.org/10.1080/03050060500317620

Karam, F. J., Monaghan, C., & Yoder, P. J. (2017). 'The students do not know why they are here': Education decision-making for Syrian refugees. *Globalisation, Societies and Education*, *15*(4), 448–463. https://doi.org/10.1080/14767724.2016.1222895

Kelly, A. V. (2004). *The Curriculum: Theory and Practice* (5th ed.). SAGE Publications Ltd.

Kerr, J. F. (Ed.). (1968). *Changing the Curriculum*. University of London Press.

Liberty, K. (2018). How research is helping our children after the earthquakes. *Christchurch Health Precinct*. https://www.healthprecinct.org.nz/news/stories/how-research-is-helping-our-children-after-the-earthquakes/

Lidher, S., McIntosh, M., & Alexander, C. (2020). Our migration story: History, the national curriculum, and re-narrating the British nation. *Journal of Ethnic and Migration Studies*, *0*(0), 1–17. https://doi.org/10.1080/1369183X.2020.1812279

López Fesser, M. (2017, April 20). *In Cuba, getting children back into classrooms after Hurricane Matthew*. In Cuba, Getting Children Back into Classrooms after Hurricane Matthew. https://www.unicef.org/stories/cuba-getting-children-back-classrooms-after-hurricane-matthew

McLaughlin, M. (2020, July 28). Coronavirus in Scotland: Mental health will be main subject in lighter school lessons. *The Times*. https://www.thetimes.co.uk/article/coronavirus-in-scotland-mental-health-will-be-main-subject-in-lighter-school-lessons-dpsqtsnx2

Mendizabal, E., & Hearn, S. (2011). *Inter-Agency Network for Education in Emergencies: A community of practice, a catalyst for change*. UNESCO IIEP. https://inee.org/sites/default/files/resources/INEE_a_community_of_practice_a_catalyst_for_change__2011.pdf

Moos, R. H., & Moos, B. S. (2006). Rates and predictors of relapse after natural and treated remission from alcohol use disorders. *Addiction (Abingdon, England)*, *101*(2), 212–222. https://doi.org/10.1111/j.1360-0443.2006.01310.x

Northern Ireland Department of Education. (2020). *Curriculum Planning 2020/21: Circular 2020/06 -*. Northern Ireland Department of Education. https://www.education-ni.gov.uk/publications/circular-202006-curriculum-planning-202021

Northern Ireland Department of Education. (2020). *Education Restart—Information and Guidance for Educational Settings*. Northern Ireland Department of Education. https://www.education-ni.gov.uk/education-restart-information-and-guidance-educational-settings

Obura, A. (2003). *Never Again: Educational Reconstruction in Rwanda*. UNESCO IIEP. https://www.iiep.unesco.org/en/publication/never-again-educational-reconstruction-rwanda

Ofsted. (2020). *COVID-19 series: Briefing on schools, September 2020*. Ofsted. https://www.gov.uk/government/publications/covid-19-series-briefing-on-schools-september-2020

Ólafsson, D. (2013). Scribal Communities in Iceland: The Case of Sighvatur Grímsson. In A. Kuismin & M. J. Driscoll (Eds.), *White Field, Black Seeds: Nordic Literacy Practices in the Long Nineteenth Century* (pp. 40–49). SKS.

Pausigere, P. (2009). *Curriculum Development in an Urban Refugee Centre in South Africa*. University of Witwatersrand. https://inee.org/resources/curriculum-development-urban-refugee-centre-south-africa

Porter, A. C., & Smithson, J. L. (2001). *Defining, Developing, and Using Curriculum Indicators* (CPRE Research Reports. RR-048). University of Pennsylvania Graduate School of Education. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1058&context=cpre_researchreports

Pozzetta, G. E., & Mormino, G. R. (1998). The Reader and the Worker: “Los Lectores” and the Culture of Cigarmaking in Cuba and Florida. *International Labor and Working-Class History*, *54*, 1–18. JSTOR.

Priestley, M. (2019, May). Curriculum: Concepts and approaches. *Impact: Journal of the Chartered College of Teaching*, *6*. https://impact.chartered.college/article/curriculum-concepts-approaches/

Ragnarsdóttir, G., & Jóhannesson, I. Á. (2014). Curriculum, crisis and the work and well-being of Icelandic upper secondary school teachers. *Education Inquiry*, *5*(1), 24045. https://doi.org/10.3402/edui.v5.24045

Reid, N., Khan, B., Soklaridis, S., Kozloff, N., Brown, R., & Stergiopoulos, V. (2020). Mechanisms of change and participant outcomes in a Recovery Education Centre for individuals transitioning from homelessness: A qualitative evaluation. *BMC Public Health, 20*(1), 497. https://doi.org/10.1186/s12889-020-08614-8

Reimers, F. M., & Schleicher, A. (2020). *A framework to guide an education response to the COVID-19 Pandemic of 2020*. Organisation for Economic Co-operation and Development (OECD). https://www.aforges.org/wp-content/uploads/2020/04/framework.pdf

Scottish Government/Riaghaltas na h-Alba. (2008). *Building the Curriculum 3: A framework for learning and teaching*. Scottish Executive. https://education.gov.scot/education-scotland/scottish-education-system/policy-for-scottish-education/policy-drivers/cfe-building-from-the-statement-appendix-incl-btc1-5/building-the-curriculum/

Scottish Government/Riaghaltas na h-Alba. (2020). *Coronavirus (COVID-19): Curriculum for Excellence in the Recovery Phase—Gov.scot*. Scottish Executive. https://www.gov.scot/publications/coronavirus-covid-19-curriculum-for-excellence-in-the-recovery-phase/

Shah, R. (2009, September 15). *Timor-Leste ten years on: Reconstructing curriculum for the future*. 10th UKFIET Conference on Education and Development, Oxford, England. https://inee.org/sites/default/files/resources/UKFIET_2009_ARTICLE.pdf

Sherwood, H. (2020, December 6). Music education in UK schools devastated by pandemic, survey finds. *The Guardian*. http://www.theguardian.com/education/2020/dec/06/music-education-in-uk-schools-devastated-by-pandemic-survey-finds

Sinclair, M. (2002). *Planning Education in and after Emergencies*. UNESCO IIEP. https://www.iiep.unesco.org/en/publication/planning-education-and-after-emergencies

Smith, M. (2013). *UNICEF Back to School Guide*. UNICEF. https://inee.org/sites/default/files/resources/UNICEF_Back_To_School_Guide_2013.pdf

Stabback, P. (2016). *What Makes a Quality Curriculum?* (IBE/2016/WP/CD/02; Current and Critical Issues in Curriculum and Learning). International Bureau of Education/UNESCO. http://unesdoc.unesco.org/images/0024/002439/243975e.pdf

Swiss National Science Foundation. (2017, January 19). School curricula are a reflection of society's expectations. *ScienceDaily*. www.sciencedaily.com/releases/2017/01/170119083549.htm

The Inter-agency Accelerated Learning Working Group (AEWG). (2020). *COVID-19 Pathways for the Return to Learning: Guidance on Condensing a Curriculum*. UNHCR. https://inee.org/sites/default/files/resources/AEWG_COVID19_Pathways%20for%20the%20Return%20to%20Learning-screen_1.pdf

Tinajero, A. (2010). *El Lector A History of the Cigar Factory Reader*. University of Texas Press. https://doi.org/10.7560/721753

UNESCO. (2017). *Education in Emergencies: Preparedness, Response, Recovery*. http://www.unesco.org/new/en/unesco/themes/pcpd/education-in-emergencies/0

UNESCO. (2020, March 4). *Education: From Disruption to Recovery*. UNESCO. https://en.unesco.org/covid19/educationresponse

UNHCR. (2012). *UNHCR Education Strategy 2012-2016*. United Nations High Commissioner for Refugees. https://www.unhcr.org/protection/operations/5149ba349/unhcr-education-strategy-2012-2016.html

UNHCR. (2019a). *Stepping up: Refugee education in crisis - 2019 Report - UNHCR*. Stepping Up: Refugee Education in Crisis - UNHCR. https://www.unhcr.org/steppingup/

UNHCR. (2019b). *Statistical Yearbooks: Figures at a Glance*. United Nations High Commissioner for Refugees. https://www.unhcr.org/figures-at-a-glance.html

UNICEF. (2000). *Curriculum Report Card* [Working Paper Series]. UNICEF. https://inee.org/sites/default/files/resources/UNICEF_2000_Curriculum_report_card.PDF

UNICEF. (2020). *The Learning Passport: Transforming Societies Through Education*. Introduction to The Learning Passport. https://www.learningpassport.org

VERBI Software. (2021). *MAXQDA 2022*. https://maxqda.com

Welsh Government. (2020a). *Guidance on learning in schools and settings from the autumn term*. Welsh Government. https://gov.wales/sites/default/files/publications/2020-07/guidance-on-learning-in-schools-and-settings-from-the-autumn-term.pdf

Welsh Government. (2020b, January). *Introduction to Curriculum for Wales guidance*. https://hwb.gov.wales/curriculum-for-wales/introduction/

Winter, C. (2018). Disrupting colonial discourses in the geography curriculum during the introduction of British Values policy in schools. *Journal of Curriculum Studies*, *50*(4), 456–475. https://doi.org/10.1080/00220272.2018.1428366

What's in a name? Are surnames derived from trades and occupations associated with lower GCSE scores?

Joanna Williamson and Tom Bramley (Research Division)

Introduction

For many readers, the image conjured up by a character called Mr or Miss Darcy will be different from the image associated with a Mr or Miss Tinker. In Thomas Hardy's social-realist novels, similarly, plots can plausibly turn on the distinction implied by d'Urberville as opposed to Durbeyfield. These expectations and associations reflect socio-cultural knowledge of surname origins, even if we understand it only implicitly. In England, Darcy and Turberville are examples of a "distinctive class of surnames" belonging to Norman, Breton and Flemish estate-owners who arrived in Britain with the Norman Conquest (Clark & Cummins, 2014, p. 525). The Darcy and Turberville families, along with others such as the Montgomery and Mandeville families, were major landowners in the Domesday Book. The hypothetical Tinker family, meanwhile, have a surname derived from the occupation of mending pots and pans.

Educational outcomes in England today vary according to socio-economic advantage: there are persistent associations between measures of educational attainment and socio-economic indicators such as parental education level and entitlement to free school meals (Sutherland et al., 2015). Research literature on the history of names, meanwhile, confirms that surnames in England – as well as many other countries – were highly socially stratified in their origins (Hanks & Parkin, 2016). The above facts prompted us to wonder whether the educational achievements of school students in England (as captured, for instance, in GCSE results) might show variation according to surname or family name origin.

Economic historians have used surnames to link cross-sectional data on socio-economic status (e.g., enrolment lists for elite universities from different centuries, as used by Clark & Cummins, 2014), and thereby measure long-term social mobility. What was not clear, was whether the origin of a surname itself (i.e., without linking it to a separate index of socio-economic status) would carry information to the extent that it would be reflected in today's GCSE results. To the best of our knowledge, this had not previously been researched. Our hypothesis was that surnames with an occupational origin would be associated with slightly lower average GCSE scores than surnames of other origins. While we know that a surname is not an empirical measure of an individual's socio-economic position, our hypothesis was that in aggregate, the educational outcomes of a group defined in this way might still reflect past social history.

Surname origins

Across the world, there exist three main systems of naming. In a binomial system, individuals are known by a given name (or sometimes several) together with a family name or surname that is inherited and subsequently passed on between generations. In patronymic naming systems, given names are instead accompanied by a name that describes parentage, for example Jakobsson or Jakobsdóttir for the son or daughter of someone named Jakob. In the Arabic naming system, meanwhile, an individual's name consists of up to five elements: besides a given name (*ism*), other possible elements include names with nickname, patronymic and locative meanings[1].

In most European countries, a binomial naming system was established between the twelfth and fourteenth centuries, as non-hereditary names began to be fixed within families and handed down from generation to generation. In England specifically, it was rare to see individuals recorded with more than one name prior to the Domesday Book of 1086. As in other European countries, non-hereditary "by-names" appeared before surnames, which added descriptive details to distinguish an individual from others with the same given name. These descriptive names fell into four main categories: reference to a trade or occupation, reference to a person's geographical location or origin, description of a relationship to another person (e.g., patronymic names), or reference to some physical characteristic or behaviour (Hanks & Parkin, 2016, p. 3). These four categories provide the most common classification for surname origin in England today (Table 1).

1 For further details of naming systems and their history, see Hanks & Parkin (2016). On the Arabic naming system in particular, Hanks and Parkin note that individuals moving from Arabic-speaking societies to countries using a binomial naming system have adopted different elements of the possible five as their surname.

Table 1: Typology of surnames in Britain and Ireland.

Surname origin	Description	Sub-types	Examples	Less obvious examples
Relationship	Reference to a relationship to another person	Patronymics	Peterson, Michaels	Brightman
Locative	Reference to a person's geographical location or origin	Topographical, toponymic	Churchill, Oppenheimer	Dubois
Occupational	Reference to a trade or occupation	Other	Archer, Fowler	Palfreyman
		Status	Laird	Pasha, Villain
Nickname	Reference to a physical characteristic or behaviour	–	King	Mordaunt

The "less obvious examples"[2] in Table 1 include surnames whose origins require a little more explanation than others. The name "Brightman" indicates a relationship to someone whose given name involved the Old English stem word "beorht" (meaning "bright"), while "Dubois" is a locative name translating literally to "of the woods", and "Palfreyman" refers to the occupation of maintaining saddle horses. The surname "Pasha" derives from the (high) rank of this name within the Ottoman empire, and, at the opposite extreme, "Villain" derives from the Anglo-Norman word *villein* (meaning *serf*). The name "Mordaunt", meanwhile, originated as a Norman nickname for someone with a sharp tongue.

Although "status" surnames are usually considered a subset of occupational surnames, they can reflect status in different ways. A surname that appears to describe a high-status role is more likely to indicate a servant to the high-status individual, or a nickname based on personal qualities. For instance, "Baron" is listed in the *Oxford Dictionary of Family Names in Britain and Ireland* as both a nickname and status name, while "Knight" is listed as a status name, occupational name, nickname and relationship name (Hanks et al., 2016). Similar explanations apply in other languages – in German, the surnames Kaiser (emperor) and König (King) originated as nicknames and indications of subordinate roles (Silberzahn & Uhlmann, 2013). As a broader point, status-origin names illustrate why research on the origin and history of names is important, even when the semantic meaning of the name seems unambiguous.

The social stratification of surnames in England included both which surnames were carried by which individuals, but also whether an individual had a surname at all (Hanks & Parkin, 2016, p. 4). While some wealthy landowners arrived in England with hereditary surnames, or adopted the use of hereditary surnames soon after the Norman conquest, others in society were still known by bynames many centuries later. Elsewhere in Britain and Ireland, research has identified varying patterns of surname establishment and development. In Ireland, surnames were established early and there has been "considerable exchange of surnames

2 All examples in Table 1 (and their classifications and explanations) come from the *Oxford Dictionary of Family Names in Britain and Ireland* (Hanks et al., 2016).

between Britain and Ireland for almost a millennium" (p. 6). In Scots-speaking regions, meanwhile, surnames began to be used at a similar date as in England, but their development towards widespread use was slower (p. 7). In Wales, the development of hereditary surnames occurred much later than in England, Ireland and Scotland – hereditary surnames were rare in Wales as late as 1500.

Surnames in research

Administrative records provide highly extensive (sometimes population-wide) data sets linking surnames to marriages and births, residency, education and ownership. In combination with their heritability, this makes surnames a valuable information source, and surnames have consequently been used by researchers in a wide range of disciplines, including genetic and demographic studies (e.g., Relethford, 1992), geography (e.g., Longley et al., 2007), sociology (e.g., Jackson, 2009), and economics and development studies (e.g., Dasar, 2019).

Economic historians have demonstrated that surnames can offer an innovative source of data for researching social mobility, not only in England but countries as diverse as Spain, Sweden, Chile, China and Korea (Clark & Cummins, 2014; Clark et al., 2015; Guell et al., 2014)[3]. In this research, surnames have been used to trace families over multiple generations, but surnames themselves have not been treated as informative. Guell et al. (2014, p. 694) emphasise this in quite strong terms, explaining that surnames are "intrinsically irrelevant" except for the fact that "they get passed from one generation to the next, alongside other characteristics that *do* matter". The argument for this is that most surnames tend towards more even distribution across social strata over time. Among long-established surnames in England, those that were commonly found in the population by 1800 were by that point associated with average levels of social status (Clark & Cummins, 2014, p. 525). Guell et al. state firmly that "We cannot learn anything from the name Smith" (p. 695). In our research, we wondered whether we could in fact learn something from the name Smith – or at least, detect differences in the GCSE outcomes of the group of students that includes Smiths, Tinkers, Bakers and Butchers, relative to students with other surnames.

Research that is concerned with the information captured by surnames themselves is rare. An example is the research by Voracek et al. (2015) into the relative physical strengths of men with the surnames Tailor and Smith. This research aimed to replicate earlier work by Bäumler (1980), who put forward a "genetic-social hypothesis" for the association of different body types with certain occupational surnames. The logic of this hypothesis was that (1) both surnames and important physical characteristics are inherited, (2) many trades (including blacksmiths) were historically organised around guilds and showed high levels of within-group marriage and apprenticeships, and (3) physical characteristics that are useful or prerequisite for a given trade will be selected for in those joining the trade. In

3 By making use of surnames that are relatively rare within a society, researchers have been able to track the social status of families across far longer time periods than in conventional studies on social mobility, which have typically been restricted to studying adjacent generations (parent–child relationships, or at most grandparent–parent–child relationships).

combination, these factors point to "above-chance preservation of any heritable traits within lineages", and might have resulted in "discernible physical differences" between men in contemporary society with the surnames Tailor and Smith (Voracek et al., 2015, p. 2).

Earlier studies found that Smiths were indeed heavier than Tailors, considered themselves more suited for activities requiring strength, and were overrepresented in sports requiring strength (Bäumler, 1980; Stemmler & Bäumler, 2003). The two new studies reported by Voracek et al. (2015) gave a mixed picture: Smiths did rate themselves more highly in strength-related activities, and Tailors rated themselves more highly in dexterity-related activities, but the effects were small and only the "Smith" effect significant. The findings also showed an increasing prevalence of Smiths from "light-stature to medium-stature to heavy-stature sports" (p. 8), but no pattern in the prevalence of Tailors in different sports, and no differences between Tailors and Smiths in their basic physical characteristics.

These findings could be explained by adaptations to Bäumler's hypothesis: for example, a stronger occupational selection effect for blacksmiths than for tailors, or fewer opportunities for assortative mating for tailors. Voracek et al. (2015) point out that the findings could also be explained by an entirely different explanatory mechanism, namely the psychological effect of implicit egotism. Implicit egotism is the idea that people have an (unconscious) preference for people, places and things that they associate with themselves, and in this case would mean that "Smiths, merely because of their surname, would feel more inclined to weight-train" and take part in strength-related sports (p. 9). Previous studies have shown that people are more likely to undertake careers and move to locations that resemble their surname. For example, Canadians with surnames beginning Tor-, Cal- and Win- are disproportionately highly represented in Toronto, Calgary and Winnipeg respectively (Pelham et al., 2002; Pelham et al., 2003).

While implicit egotism is about the automatic associations that individuals make in relation to themselves, surnames may also have impact through the associations made by others. A study by Silberzahn and Uhlmann (2013), for example, showed that Germans with noble-sounding names such as König (King) are over-represented in management positions, compared to Germans with surnames that refer to ordinary occupations such as Becker/Bäcker (baker) or no social role. Silberzahn and Uhlmann (2013) hypothesise that this is due to associative cognition on the part of others: the high status associated with the noble-sounding name "may implicitly spill over to its bearer, influencing the status accorded to that person and consequential life outcomes" (p. 2441). For instance, the status may lead to more positive interpretations of traits and performance in the workplace. The influence of a noble-sounding name on an individual's self-perception may also be a factor – perhaps by encouraging them to pursue high-status roles.

A final area of surname research to consider is on "alphabetism", that is, the impact of having a surname that starts with a letter ranked later in the alphabet. Cauley and Zax (2018), for example, showed that surname initials ranked further from the beginning of the alphabet were associated with lower educational attainment, as well as other poor life outcomes, for men (women were excluded from the study).

Names and social biases

Name effects have been more extensively researched in relation to negative social discrimination based on race, class and culture. English-language studies in this area have focused more frequently on given names than surnames, but it is worth outlining some notable findings.

Experimental studies have used manipulation of names on student work or application letters to test the effect of (inferred) race, class or culture on the judgements made about an individual by others – particularly teachers and employers. As summarised by Fryer and Levitt (2004, p. 771), these studies have repeatedly shown that "resumes with traditional names are substantially more likely to lead to job interviews than are identical resumes with distinctively minority-sounding names". An important contribution by Fryer and Levitt was to show that the association between a distinctively minority-sounding name and life outcomes is no longer found once socio-economic circumstances at birth are controlled for.

Also in the US context, Figlio (2005) has argued that teacher expectations are sensitive to the perceived socio-economic status as well as racial status of school students' given names, with effects on school achievement. Students whose given names are associated with low socio-economic status achieve lower test results than their siblings and fellow students whose names are less class-identifiable (pp. 21–22), and this striking finding is evident even between twins. Consistent with Fryer and Levitt (2004), Figlio (2005) found that the perceived class status of names had a larger effect than the perceived racial status of a name. The explanatory mechanism suggested was that teachers "may use a child's name as a signal of unobserved parental contributions to that child's education, and expect less from children with names that 'sound' like they were given by uneducated parents" (p. 1). Data on teachers' referrals to "gifted" programmes, as well as promotions of students to the next grade, were consistent with this proposed explanation.

In a UK experiment, meanwhile, Jackson (2009) found that (fictitious) job applicants whose name, school type and interests were all associated with high social status were more likely to receive a reply from employers than candidates whose equivalent characteristics were associated with low social status. Jackson notes that first names in England can offer "extremely strong signals of class origin" (examples used in the study were "Camilla Bevans-Brown" and "Donna Taylor"). No difference was found in the rate of positive replies, however, and, most interestingly, Jackson concluded that employers were responding to the signal implied by sets of characteristics, and that no individual characteristic conferred a benefit.

Educational outcomes in England

The important context underlying this research is that educational outcomes in England (as well as in many other countries) are known to vary by socio-economic status (SES) (e.g., DfE, 2019, pp. 8–9). In an investigation of different proxies for socio-economic status, Sutherland et al. (2015) showed that after controlling for other characteristics, free school meal (FSM) eligibility in the preceding five years "equates to the difference between a pupil gaining one grade better across seven GCSEs (e.g., moving from a C to a B) and two grades better on an eighth GCSE" (p. 8), and that parental occupation, parental education, and other household measures were even better predictors of a pupil's educational outcomes.

Educational outcomes also vary by ethnic group (e.g., DfE, 2019, p. 12), by gender (Bramley et al., 2015) and by age relative to others in the same cohort (Benton, 2014). Further, the gaps in attainment between low-SES (e.g., FSM) pupils and other pupils also vary by subgroup (DfE, 2022). For example, the attainment gap by FSM status is much larger among white pupils than among those in other ethnic groups. Low socio-economic groups include disproportionate numbers of minority ethnic students, but minority ethnic students make greater progression during secondary education than white British students, after accounting for prior attainment (Leckie & Goldstein, 2019; Wilson et al., 2009). While differences can be explained to some extent by language, cultural attitudes towards education and qualifications are also hypothesised to play a role (Hoffmann, 2018; Wilson et al., 2009).

Data and method

The research was designed to test the simple hypothesis that average GCSE results would be lower among candidates whose surname originated as an occupation than among other candidates. To give context, we were also interested in how the observed difference (if any) compared to the difference in GCSE scores by gender, and the difference in GCSE scores by birth month. Finally, we decided to look at whether the GCSE scores showed evidence of alphabetism.

Data

We obtained all results in GCSE Mathematics (A*–G) and GCSE English (A*–G) from the awarding body OCR, for the years 2012–17. We retained only those candidates for whom we had at least one GCSE Mathematics grade and at least one GCSE English grade, taken at the usual age[4] of 15–17. In this article "mean GCSE score" refers to the average of these two GCSE grades (after converting the A*–G letter grades to numbers 8–1, and taking the best result if candidates had more than one grade in either subject).

The purpose of using GCSE English and Mathematics was to obtain a large data set on educational attainment that was as free as possible from subject selection effects (since GCSE Mathematics and English are taken by almost all 16-year-olds in England). The final data set contained a mean GCSE score for just under 21 000 unique candidates. The data set also included candidate surname, gender and date of birth.

4 We excluded very early entry candidates and results from learners aged 18 or over to avoid age effects as far as possible. Restricting to candidates aged 16 exactly, however, would have reduced the available data too far (by around 25 per cent), since sitting GCSE English and/or Maths one year early, or re-sitting one of these in Year 12, was fairly common during 2012–17.

For surname classification, we used the database underpinning the *Oxford Dictionary of Family Names in Britain and Ireland* (Hanks et al., 2016), from here on abbreviated to FaNBI. This database is an up-to-date and research-based authority on UK surnames[5], listing all surnames held by 100 or more individuals in 1997. In addition, FaNBI lists surnames that have appeared in other British surname dictionaries, and "established names" (those found in both the 1881 census and 1997 electoral rolls) (Hanks & Parkin, 2016, p. 12).

As shown by the examples in Table 1, established UK surnames derive from a range of languages and cultures, reflecting many centuries of immigration history. The FaNBI database lists information on the language and/or cultural origin of listed surnames, where known. The FaNBI database also includes classification of each surname's origins according to the typology shown in Table 1. For the purposes of our analysis, we decided to record status-origin surnames as a category in their own right (rather than a subset of occupational surnames). This was for two reasons: firstly, the explanation of status-origin names (in relation to the semantic meaning of the name) is quite different from the explanation of other occupational names, and secondly, the literature suggests that status-origin names could be affected by the associative cognition effects described by Silberzahn and Uhlmann (2013). As for the language and culture field in FaNBI, the surname typology field for each surname could be blank, list one surname type, or list several types.

Data preparation and classification

We created an indicator variable for "Occupational" surname origin. Candidate surnames listed in FaNBI were flagged as "Occupational" if "Occupational" appeared in the typology list for the name, or if the surname was included in FaNBI as a variant of a listed surname which itself had "Occupational" in the typology list. Indicators for "Status", "Relationship", "Locative" and "Nickname" surname origins were created in the same way. It is important to emphasise that these surname origin indicators were a set of five independent binary indicator variables, rather than a classification variable, and that the same surname could be flagged by multiple indicators. The reason for this was that, as noted above, surnames in the FaNBI database could have zero, one, or several different origins listed in the surname typology field.

We also created an indicator variable to record whether a name derived from a British or Irish language, narrowly defined: surnames were flagged if the language/culture field in FaNBI included "English", "Welsh", "Scottish", "Irish", "Manx", "Cornish", or "Norman". Because socio-economic status, educational progression and educational outcomes can vary across ethnic and cultural groups, we were concerned about the possibility of conflating surname effects with ethnic and

5 Rather than, for instance, a re-publication, new edition or amalgamation of previous surname dictionaries. For a useful overview of the scholarship on surnames in Britain and Ireland, see Hanks and Parkin (2016). The project to create the FaNBI database and dictionary was initiated by Oxford University Press and the Arts and Humanities Research Council of Britain. The research was led by academics at the University of the West of England, in association with database experts at Brno University in the Czech Republic.

cultural group effects. Without data on candidate ethnicity or culture, we could not directly control for this, so the purpose of the British/Irish language indicator was to offer an imperfect proxy. The indicator enabled us to compare results of analyses for all FaNBI-listed surnames with results for linguistically British/Irish surnames only.

Multi-part names (e.g., double-barrelled) were split, and we obtained classification details for each part. The exception to this was multi-part names including common prefix words such as "De la…", "Al…" or "Von…". In some of these cases, the entire multi-part name was listed in FaNBI as one name, allowing easy classification. If the entire name was not listed in FaNBI, we attempted a re-classification after discarding the prefix words.

The surname origin indicators were then applied to candidates' surnames: candidates were flagged if their entire surname or part (if a multi-part name) met the conditions for inclusion. So, for instance, a candidate with the surname Carter-Khan would have been flagged as having an occupational-origin name.

To make sure that no individuals would be identified, we removed given names and school names from the GCSE results data set. In the results, we do not mention or list any surnames belonging to fewer than 10 candidates in the results data set.

Analyses

We used descriptive statistics to explore the proportion of candidates in the results data set whose names we could categorise using the FaNBI database, and the surname types of those whose names were listed. We also described how average mean GCSE scores varied by surname origin, birth month and gender. We then used multilevel linear models to investigate the relationship between mean GCSE score and surname origin.

To investigate the "First letter" alphabetism hypothesis, we calculated the correlation between mean GCSE scores and surname initial letter (as a numerical rank).

Results

The results data set contained 8681 different candidate surnames[6], of which just over 80 per cent were listed in FaNBI. The candidate surnames not listed in FaNBI tended to account for smaller numbers of candidates (i.e., they were the less common surnames), so overall, over 19 000 candidates (more than 91 per cent of those in the results data set) had a surname that was listed in FaNBI.

Table 2 shows the numbers of candidates flagged by each of the surname indicators. The surname origin indicators are not mutually exclusive categories, and some candidates had names flagged by multiple indicators. Of those students whose surname appeared in FaNBI, 92 per cent had a surname of British/Irish language origin (henceforth, abbreviated to BIL). A high proportion had a surname that was a relationship name (39 per cent) or locative name (40 per

6 This total was obtained from treating each unique candidate surname at face value (i.e., Clark, Clarke, Smith-Clark and Smith would be counted as four names, even though FaNBI identifies Clarke as a variant spelling of Clark).

cent), and just under 16 per cent had a surname derived from an occupation. Examples of surnames for each surname indicator can be found in the Appendix.

Table 2: Candidates identified by each indicator, among those with surnames listed in FaNBI.

Surname indicators	Percentage of candidates (out of 19 023)
British/Irish language name	92.1
Occupation name	15.9
Status name	1.8
Relationship name	38.8
Locative name	40.2
Nickname name	20.8

GCSE results by surname

We first looked at the mean GCSE scores of common occupational surnames. Figure 1 shows the average mean GCSE scores for all those occupational surnames that appeared at least 30 times in the results data set. For reference, it also shows the three surnames with the highest[7] average mean GCSE scores (Chan, Alexander and Jennings), the three surnames with the lowest average mean GCSE scores (Wheeler, Weaver and Mellor), and the average mean GCSE score for all candidates in the result data set. Unsurprisingly, the average mean GCSE scores for the commonly found occupational surnames are clustered around the average.

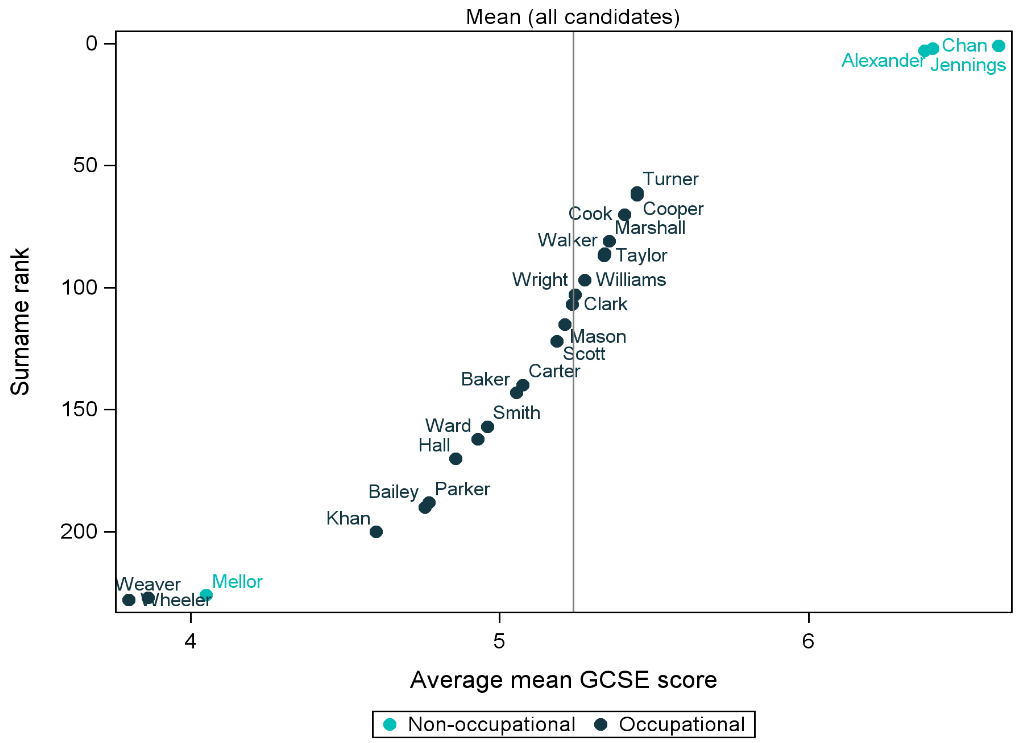

Figure 1: Average mean GCSE scores for common occupational surnames in the results data set.

7 Considering only surnames with at least 10 occurrences in the results data set. Note that this reduced the number of different surnames from over 7000 to 289.

Analysis of candidates' mean GCSE scores by surname confirmed that after accounting for known sources of variation (gender, age and school/centre), there was significant variation between different surnames[8].

We next looked at whether there were differences in the mean GCSE scores according to each surname origin indicator (separately), as shown in Figure 2. The average mean GCSE score of candidates with an occupational-origin surname was slightly lower than for those without an occupational-origin surname, in line with the research hypothesis. There was a much larger difference, however, according to whether candidates had a status-origin surname.

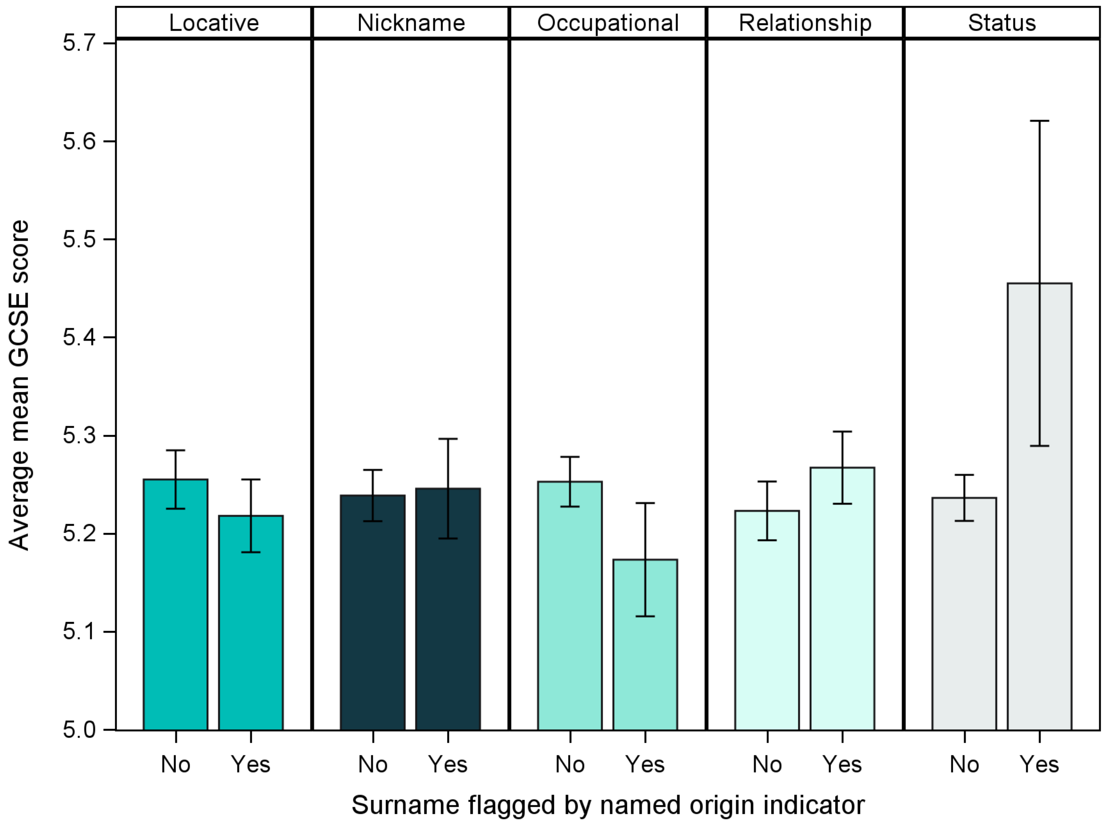

Figure 2: Differences in mean GCSE score by surname origin.

Because of the large difference in mean GCSE score according to status origin, and the fact that a surname in FaNBI could be recorded as having both an occupational and status origin, we decided to use the four-way surname classification shown in Table 3 for the remainder of the analysis. Table 3 shows that candidates with an occupational (non-status) surname had the lowest mean GCSE scores, with a mean of 5.16. At the other extreme, candidates with a status-

8 We fitted a multilevel model with gender, age and centre as fixed effects, and surname as a random effect. The variance of the random surname intercepts was small but statistically significant (0.012, p=0.038). The data set included all candidates with a surname listed in FaNBI (N=19 023), and names were analysed "as seen" (i.e., surnames listed as variants or possible variants of one another were not re-coded as the same name).

origin (non-occupational) surname had the highest mean GCSE scores, with a mean of 5.48. When the data set was restricted to consider only BIL surnames, the mean GCSE score averages for the two status-origin surname groups reduced, resulting in a smaller difference between the status-origin surname groups and the occupational (non-status) surname group.

Table 3: GCSE results by surname group.

	All candidates with surname in FaNBI			Candidates with BIL surnames		
Surname group	**N**	**N unique names**	**Mean GCSE (Std Error)**	**N**	**N unique names**	**Mean GCSE (Std Error)**
Occupational (non-status)	2879	657	5.16 (0.03)	2808	633	5.17 (0.03)
Non-status, non-occupational	15 811	6256	5.25 (0.01)	14 436	5610	5.25 (0.01)
Occupational and status	144	56	5.42 (0.13)	142	54	5.40 (0.13)
Status (non-occupational)	189	50	5.48 (0.11)	141	39	5.33 (0.13)
Total	*19 023*	*7019*		*17 527*	*6336*	

Context

To contextualise the differences in mean GCSE score by surname group, we plotted the distributions of mean GCSE scores according to birth month and gender, two other variables for which GCSE result effects are observed (Figure 3).

The average mean GCSE score for candidates with an occupational (non-status) surname was 5.16, while for candidates whose surnames had neither occupational nor status origins the average mean GCSE score was 5.25. The distributions in Figure 3 show that this difference of 0.09 GCSE points was comparable in size to the difference associated with several months' difference in birth month, and about a quarter as large as the difference in mean GCSE scores by gender.

There was a larger difference in mean GCSE scores when candidates with occupational (non-status) surnames were compared to candidates with status-origin surnames. The most extreme was the difference of 0.32 GCSE points between candidates with an occupational (non-status) surname (mean 5.16) and candidates with a status-origin (non-occupational) surname (mean 5.48). This difference was greater than the difference between September-born and August-born students, and was approximately three-quarters the size of the observed difference by gender.

Among just the BIL surnames, the largest difference by surname group was between candidates with an occupational (non-status) surname (mean GCSE 5.17) and candidates with a status-origin and occupational surname (mean GCSE 5.40), as shown in Table 3. This difference of 0.23 GCSE points was the same size as the difference between September-born and August-born students, and roughly half the size of the observed difference by gender.

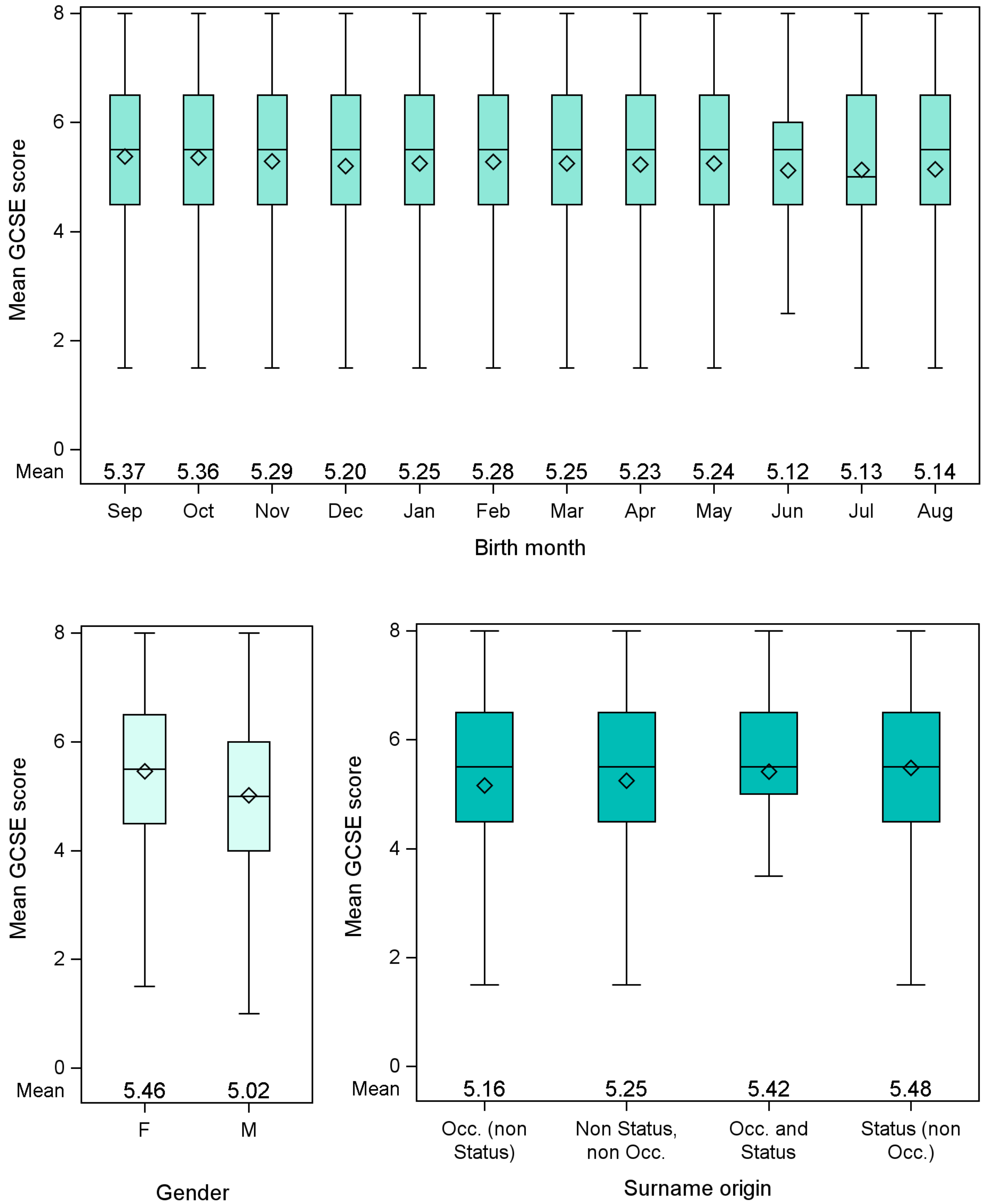

Figure 3: Distributions of mean GCSE score, by birth month, gender and surname origin.

Models

We estimated multilevel linear regression models for mean GCSE score, for candidates with a surname listed in FaNBI. A multilevel model structure was used in order to account for the clustering of students within schools.

The simplest model (Model 1) used the following structure:

> mean GCSE score = surname group + random intercept for centre.

In Model 2, we included the additional contextual variables of gender and birth month:

> mean GCSE score = surname group + gender + birth month + random intercept for centre.

In Model 1, the estimated effect of an occupational-origin (non-status) surname, in comparison with the baseline category of a non-status non-occupational surname, was -0.06 GCSE points ($p=0.028$). The estimated effect of having a status-origin (non-occupational) surname was 0.16 ($p=0.091$).

When the additional variables of gender and birth month were added (Model 2), the estimated effects associated with an occupational-origin (non-status) surname and non-status non-occupational surname changed very little from Model 1. The size of the estimated gender effect was larger than either surname effect (-0.28, $p<.0001$), while the estimated effect of birth month was smaller (-0.01, $p<.0001$).

When the models were re-estimated using only the candidates with BIL surnames (Model 3 and Model 4) the size of the effect associated with an occupational-origin (non-status) surname remained unchanged (-0.05), but there was an increase in p-value, so that the effect was no longer statistically significant at the 5 per cent level. The estimated effects for gender and birth month remained very similar to those found in Model 2.

Table 4: Estimated model parameters.

	Model 1*		Model 2*		Model 3†		Model 4†	
Effect	**Estimate (SE)**	**Pr > |t|**	**Estimate (SE)**	**Pr > |t|**	**Estimate (SE)**	**Pr > |t|**	**Estimate (SE)**	**Pr > |t|**
Intercept	4.27 (0.10)		4.52 (0.10)	<.0001	4.22 (0.10)		4.47 (0.10)	<.0001
Occupational (non-status)	-0.06 (0.03)	0.028	-0.05 (0.03)	0.031	-0.05 (0.03)	0.061	-0.05 (0.03)	0.065
Occupational and status	0.13 (0.10)	0.202	0.14 (0.10)	0.178	0.12 (0.11)	0.258	0.13 (0.10)	0.221
Status (not occupational)	0.16 (0.09)	0.091	0.17 (0.09)	0.069	0.12 (0.11)	0.261	0.13 (0.11)	0.235
[Non-status, non-occupational]	0	.	0	.	0	.	0	.
Gender M			-0.28 (0.02)	<.0001			-0.28 (0.02)	<.0001
Gender F			0	.			0	.
Birth month			-0.01 (0.00)	<.0001			-0.02 (0.00)	<.0001

*Model estimated using all candidates in results data set with a surname listed in FaNBI, and both gender and birth month data available (n=19 022)

†Model estimated using only candidates with British/Irish language surname, and both gender and birth month data available (n=17 526)

Alphabetism

Figure 4 shows the average mean GCSE score for surnames beginning with each letter, together with 95 per cent confidence intervals (results for X are not plotted, as there were fewer than 10 candidates with X-surnames). In contrast to the alphabetism hypothesis, mean GCSE scores did not decrease for surnames with initials farther from the beginning of the alphabet. Calculating correlations confirmed that there was no association between mean GCSE score and the initial letters of candidate surnames ($r = 0.01$, $p = 0.193$, $N = 19\,023$). This remained the case when the analysis was repeated for BIL surnames only ($r = 0.0003$, $p = 0.973$, $N=17\,527$).

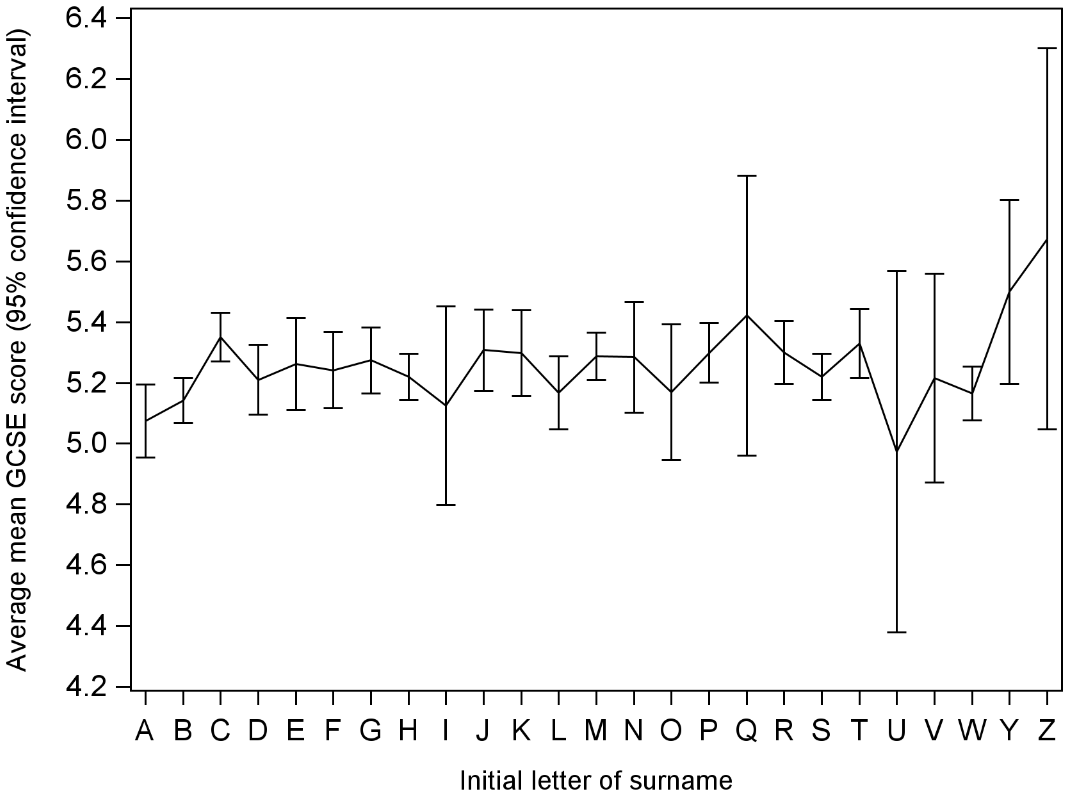

Figure 4: Average mean GCSE scores (with 95 per cent confidence intervals) by initial surname letter.

Discussion

The results of this simple study showed that the mean GCSE scores of candidates with occupational surnames were slightly lower than the mean GCSE scores of candidates with other surnames. This is in line with the research hypothesis, but the difference in GCSE attainment was not large: it was a similar size to the average difference expected between candidates half a year apart in age, and much smaller than the well-known “gap” between male and female GCSE candidates.

The size of the estimated occupational surname effect was consistent across all models estimated. The associated p-value, however, increased when the data set was restricted to candidate surnames from British or Irish languages, moving from

~0.03 to ~0.06 and consequently over the conventional threshold for statistical significance. Taken together, the regression model outcomes indicate a result that is unlikely to have arisen purely by chance. At the same time, we emphasise that the British/Irish language indicator was an imperfect proxy for investigating the impact of ethnic or cultural group differences on the surname investigation, and further research, using data on candidate ethnicity and ideally family immigration background, would be needed in order to better understand this.

This study identified a small negative effect associated with occupational non-status surnames. The explanation for this effect was beyond the scope of the current research, but surname mechanisms proposed in the literature include the psychological (e.g., implicit egotism, associative cognition), sociological (e.g., reading surnames as information signals about social class) and socio-genetic (e.g., Bäumler's (1980) "genetic-social" explanation for Tailor–Smith differences). This study hoped to offer a novel look at educational attainment and social inequalities. Ultimately, the findings are a reminder that these are highly complex matters, and that caution is needed to avoid over-interpreting small differences.

Acknowledgements

We would like to thank Professors Richard Coates, Patrick Hanks and Peter McClure and Dr Adam Rambousek for their generous assistance accessing the *Oxford Dictionary of Family Names in Britain and Ireland* (FaNBI) data base.

References

Bäumler, G. (1980). Differences in physique in men called 'Smith' or 'Tailor' considered as results of a genetic effect dating back over several centuries. *Personal and Individual Differences, 1*, 308–310.

Benton, T. (2014). *Should we age-standardise GCSEs?* Cambridge Assessment Research Report. https://www.cambridgeassessment.org.uk/Images/285155-should-we-age-standardise-gcses-.pdf

Bramley, T., Vidal Rodeiro, C., & Vitello, S. (2015). *Gender differences in GCSE*. Cambridge Assessment Research Report. https://www.cambridgeassessment.org.uk/Images/gender-differences-in-gcse.pdf

Cauley, A. S., & Zax, J. S. (2018). *Alphabetism: The effects of surname initial and the cost of being otherwise undistinguished*. SSRN. https://ssrn.com/abstract=3272556 or http://dx.doi.org/10.2139/ssrn.3272556

Clark, G., & Cummins, N. (2014, Dec). Surnames and social mobility in England, 1170-2012. *Human Nature, 25*(4), 517–537. https://doi.org/10.1007/s12110-014-9219-y

Clark, G., Cummins, N., Hao, Y., & Vidal, D. D. (2015). Surnames: A new source for the history of social mobility. *Explorations in Economic History, 55*, 3–24. https://doi.org/10.1016/j.eeh.2014.12.002

Dasar, B. N. S. (2019). A methodology to track economic mobility of castes in south Indian village, circa 1866–2001. *Journal of Rural Development, 38*(4), 580–6074.

DfE. (2019). *Key stage 4 performance, 2019 (revised)*. Department for Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/863815/2019_KS4_revised_text.pdf

DfE. (2022, March 18). *GCSE results (Attainment 8)*. UK Government. https://www.ethnicity-facts-figures.service.gov.uk/education-skills-and-training/11-to-16-years-old/gcse-results-attainment-8-for-children-aged-14-to-16-key-stage-4/latest

Figlio, D. N. (2005). *Names, Expectations and the Black-White Test Score Gap* (NBER Working Paper No. w11195). https://ssrn.com/abstract=684721

Fryer, R. G., & Levitt, S. D. (2004). The Causes and Consequences of Distinctively Black Names. *The Quarterly Journal of Economics, 119*(3), 767–805.

Guell, M., Rodriguez Mora, J. V., & Telmer, C. I. (2014). The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating. *The Review of Economic Studies, 82*(2), 693–735. https://doi.org/10.1093/restud/rdu041

Hanks, P., Coates, R., & McClure, P. (Eds.). (2016). *The Oxford Dictionary of Family Names in Britain and Ireland*. Oxford University Press.

Hanks, P., & Parkin, H. (2016). Family Names. In C. Hough (Ed.), *The Oxford Handbook of Names and Naming*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199656431.013.9

Hoffmann, N. I. (2018). Cognitive achievement of children of immigrants: Evidence from the Millennium Cohort Study and the 1970 British Cohort Study. *British Educational Research Journal, 44*(6), 1005–1028. https://doi.org/10.1002/berj.3476

Jackson, M. (2009, Dec). Disadvantaged through discrimination? The role of employers in social stratification. *British Journal of Sociology, 60*(4), 669–692. https://doi.org/10.1111/j.1468-4446.2009.01270.x

Leckie, G., & Goldstein, H. (2019). The importance of adjusting for pupil background in school value added models: A study of Progress 8 and school accountability in England. *British Educational Research Journal, 45*(3), 518–537. https://doi.org/10.1002/berj.3511

Longley, P. A., Webber, R., & Lloyd, D. (2007). The Quantitative Analysis of Family Names: Historic Migration and the Present Day Neighborhood Structure of Middlesbrough, United Kingdom. *Annals of the Association of American Geographers, 97*(1), 31–48. https://doi.org/10.1111/j.1467-8306.2007.00522.x

Pelham, B. W., Carvallo, M., DeHart, T., & Jones, J. T. (2003, Nov). Assessing the validity of implicit egotism: a reply to Gallucci (2003). *Journal of Personality and Social Psychology, 85*(5), 800–807. https://doi.org/10.1037/0022-3514.85.5.800

Pelham, B. W., Mirenberg, M. C., & Jones, J. T. (2002). Why Susie sells seashells by the seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology, 82*(4), 469–487. https://doi.org/10.1037/0022-3514.82.4.469

Relethford, J. (1992). Analysis of Marital Structure in Massachusetts Using Repeating Pairs of Surnames. *Human Biology, 64*(1), 25–33.

Silberzahn, R., & Uhlmann, E. L. (2013, Dec). It pays to be Herr Kaiser: Germans with noble-sounding surnames more often work as managers than as employees. *Psychological Science, 24*(12), 2437–2444. https://doi.org/10.1177/0956797613494851

Stemmler, M., & Bäumler, G. (2003). Testing the statistical association between family names and success in certain athletic disciplines in men called 'Smith' or 'Tailor'. *Psychological Test and Assessment Modeling, 45*(2), 254.

Sutherland, A., Ilie, S., Vignoles, A., RAND Europe, & the University of Cambridge. (2015). *Factors associated with achievement: key stage 4 (DFE- RR407)*. Department for Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/473673/RR407_-_Factors_associated_with_achievement_-_key_stage_4.pdf

Voracek, M., Rieder, S., Stieger, S., & Swami, V. (2015). What's in a Surname? Physique, Aptitude, and Sports Type Comparisons between Tailors and Smiths. *PLoS One, 10*(7), e0131795. https://doi.org/10.1371/journal.pone.0131795

Wilson, D., Burgess, S., & Briggs, A. (2009). The dynamics of school attainment of England's ethnic minorities. *Journal of Population Economics, 24*(2), 681–700. https://doi.org/10.1007/s00148-009-0269-0

Appendix

The examples in Tables 5 and 6 are taken from the FaNBI database; some but not all appeared in the results data set. Asterisks indicate names **not** listed in FaNBI as deriving from a British or Irish language (i.e., the Language/Culture field contained none of the following: “English”, “Welsh”, “Scottish”, “Irish”, “Manx”, “Cornish”, or “Norman”). The names with asterisks were therefore excluded in the analyses restricted to British/Irish language names.

Table 5: Examples of surnames with each surname indicator (not mutually exclusive).

Locative	Relationship		Occupational	Nickname	Status
Allen	Adams	Matthews	Bailey	Bell	Butler
Bailey	Allen	Mills	Baker	Brown	Chowdhury*
Bell	Anderson	Moore	Carter	Campbell	Fry
Benton	Anning	Morgan	Chamberlain	Chamberlain	Gentleman
Botham	Bell	Morris	Chapman	Cox	Guest
Bramley	Bennett	Owen	Clarke	Fry	Knight
Burgh	Brown	Pearson	Cook	Gray	Laird
Cox	Collins	Phillips	Cooper	Green	Lehmann*
Crawley	Cox	Powell	Cox	Guest	McIntosh
Darcy	Davies	Price	Fisher	Jenkins	Patel*
Fisher	Dickens	Richard	Foster	King	Prior
Graham	Edwards	Richardson	Grainger	Knight	Senior
Gray	Ellis	Robertson	Hunt	Lloyd	Stewart
Green	Evans	Robinson	Jagger	Mitchell	Tarrier
Hall	Foster	Roger	Knight	Moore	Tennant
Hill	Gibson	Russell	Marshall	Morris	Yoke
Holmes	Green	Scott	Mason	Palmer	
Jones	Griffiths	Simpson	Miller	Price	
Kelly	Harris	Swift	Parker	Prior	
Lee	Harrison	Thomas	Phillips	Reid	
Mills	Harvey	Thompson	Potter	Russell	
Moore	Hill	Watson	Prior	Senior	
Murray	Hughes	White	Smith	Shakespeare	
Newton	Jackson	Wilkinson	Stewart	Swift	
Russell	James	Williams	Taylor	Tarrier	
Shaw	Jenkins	Williamson	Tinker	Turner	
Simpson	Johnson	Wilson	Turner	White	
Smith	King		Walker	Wood	
Turner	Knight		Ward	Young	
White	Lewis		Webb		
Wood	Martin		Williams		
			Wright		

Table 6: Examples of surnames in each surname group (mutually exclusive categories).

Occupational (non-status)	Occupational and status	Status (non-occupational)	Non-status, non-occupational
Bailey	Ackerman	Agha*	Ahmed*
Baker	Batchelor	Alderman	Begum*
Carter	Fentiman	Baron	Brown
Clarke/Clark	Hackman	Bond	Davis/Davies
Cohen	Henman	Butler	Edwards
Cook	Hodgman	Chowdhury*	Evans
Cooper	Holder	Fouracre	Green
Fisher	Knight	Franklin	Harris
Fletcher	Maidman	Freeman	Hill
Foster	Master	Fry	Jackson
Harper	Monkman	Gentleman	Johnson
Hunter	Nutman	Headman	Jones
Kantor*	Paxman	Heritage	King
Mason	Parson	Laird	Lee
Parker	Prior	Le Maistre*	Lewis
Potter	Richter*	Pasha*	Martin
Slater	Sargent	Patel*	Moore
Smith	Servant	Portman	Morris
Spencer	Squire	Rabin	Roberts
Taylor	Stewart	Schultz*	Robinson
Turner	Swain	Tennant	Thomas
Walker	Swan	Vassall	Thompson
Ward	Tillman	Vavasour	White
Williams	Tubman	Villain	Wilson
Wright	Waterman	Yeoman	Wood

Research News

Lisa Bowett (Research Division)

Publications

The following reports and articles have been published since *Research Matters*, Issue 33:

Journal articles

Leech, T., Gill, T., Hughes, S., & Benton, T. (2022). The accuracy and validity of the simplified pairs method of comparative judgement in highly structured papers. *Frontiers in Education, 7* https://doi.org/10.3389/feduc.2022.803040

Walland, E., & Shaw, S. (2022). E-portfolios in teaching, learning and assessment: tensions in theory and praxis. *Technology, Pedagogy and Education*, *31*(3), 363–379 https://doi.org/10.1080/1475939X.2022.2074087

Gill, T. (2022). New research reveals just how valuable Cambridge Checkpoint assessments can be in improving results at Cambridge iGCSE. *Cambridge Outlook* issue 40

Shaw, S. D., Rushton, N., & Majewska, D. (2022). Tracing the trajectory of mathematics teaching across two contrasting educational jurisdictions: A comparison of historical and contemporary influences. *The International Education Journal: Comparative Perspectives, 21*(1), 41–60.

Editorials

Oates, T. (2022). We must not sleepwalk into ditching the national curriculum. *Schools Week*. https://schoolsweek.co.uk/we-must-not-sleepwalk-into-ditching-the-national-curriculum/

Reports

Gawedzka, G. & Gill, T (2022). *Uptake of GCSE Subjects 2018. Statistics Report Series No.125.* Cambridge University Press & Assessment.

Gawedzka, G. & Gill, T (2022). *Uptake of GCSE Subjects 2019. Statistics Report Series No.126.* Cambridge University Press & Assessment.

Gawedzka, G. & Gill, T (2022*). Provision of GCSE Subjects 2018. Statistics Report Series No.127*. Cambridge University Press & Assessment.

Gawedzka, G. & Gill, T (2022*). Provision of GCSE Subjects 2019. Statistics Report Series No.128*. Cambridge University Press & Assessment.

Gill, T. (2022). *The impact of the introduction of Progress 8 on the uptake of qualifications in English schools – an update for 2017/18 to 2019/20.* Cambridge University Press & Assessment.

Williamson, J., & Vidal Rodeiro, C. (2022). Student rank order and post-16 subject choices. Cambridge University Press & Assessment.

Conference presentations

Williamson, J. & Vidal Rodeiro, C. (2022, June 23). *Understanding post-16 subject choices: an empirical study into the role of student rank order* [Paper presentation]. Annual conference of the British Education Studies Association, Manchester.

Crisp, V., & Macinska, S. (2022, July 14). *Creating better tests: Students' views on the accessibility of different exam question design features.* [Paper presentation] 10th European Conference on Education, UCL, London, UK and online.

Kreijkes, P. (2022, July 14). *A bird's eye view on curriculum publications concerning seven countries: A bibliometric analysis* [Paper presentation]. 10th European Conference on Education, UCL, London, UK and online

Blogs

The following blogs have been published since *Research Matters*, Issue 33:

Leech, T. (2022, June 16). Value of qualifications.

Vidal Rodeiro, C. L. (2022, June 20). Cancelled GCSEs no barrier to starting post-16 study in England research shows.

Greatorex, J., & Vitello, S. (2022, January 26). What is competence? A shared interpretation of competence to support teaching, learning and assessment.

Sharing our research

We aim to make our research as widely available as possible. Listed below are links to the places where you can find our research online:

Journal papers and book chapters: https://www.cambridgeassessment.org.uk/our-research/all-published-resources/journal-papers-and-book-chapters/

Research Matters (in full and as PDFs of individual articles): https://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-matters/

Conference papers: https://www.cambridgeassessment.org.uk/our-research/all-published-resources/conference-papers/

Research reports: https://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-reports/

Data Bytes: https://www.cambridgeassessment.org.uk/our-research/data-bytes/

Statistics reports: https://www.cambridgeassessment.org.uk/our-research/all-published-resources/statistical-reports/

Blogs: https://www.cambridgeassessment.org.uk/blogs/

Insights (a platform for sharing our views and research on the big education topics that impact assessment around the globe): https://www.cambridgeassessment.org.uk/insights/

Our YouTube channel: https://www.youtube.com/channel/UCNnk0pi7n4Amd_2afMUoKGw contains Research Bytes (short presentations and commentary based on recent conference presentations), our online live debates #CamEdLive, and podcasts.

You can also learn more about our recent activities from Facebook, Instagram, LinkedIn and Twitter.

Contents / Issue 34 / Autumn 2022

Cambridge University Press & Assessment
Shaftesbury Road
Cambridge
CB2 8EA
United Kingdom

ResearchDivision@cambridge.org
www.cambridge.org

Research Matters / 34
A Cambridge University Press & Assessment publication

ISSN: 1755-6031

Journal homepage: https://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-matters/

Learning loss in the Covid-19 pandemic: teachers' views on the nature and extent of loss

Matthew Carroll and Filio Constantinou (Research Division)

To cite this article: Carroll, M., & Constantinou, F. (2022). Learning loss in the Covid-19 pandemic: teachers' views on the nature and extent of loss. *Research Matters: A Cambridge University Press & Assessment publication, 34*, 6–25

To link this article: https://www.cambridgeassessment.org.uk/Images/research-matters-34-learning-loss-in-the-covid-19-pandemic.pdf

Abstract:
The Covid-19 pandemic caused unprecedented disruption to education around the world. As education systems gradually return to normal, there is a push to understand effects of the disruption. A major impact on students is "learning loss", in which attainment and progress may have fallen behind expected levels. Various efforts have been made to quantify learning loss, but to better understand it, further work, combining quantitative and qualitative approaches, is required.

Here, we sought to record teachers' views on how far behind (or ahead) their students were compared to a "typical" year, and to gather their opinions about what had been lost (or gained). To do this, we surveyed teachers in schools that work with Cambridge CEM. We received over 400 responses, spread across 38 countries and 198 schools, thus giving a broad sample of experiences.

A majority of respondents felt their students were behind expectations. 1–2 months behind was the most common estimate, but some respondents made much larger estimates of loss, while a sizeable minority thought that their students were on track or even ahead of expectations. Descriptions of the areas of loss indicated that fundamental literacy and numeracy skills had been affected, as had practical skills and general study skills. Responses also described variable impacts, both within and between groups of students.

Effects of Covid-related disruption on education are ongoing and may be felt for some time still to come. By exploring the nature and extent of learning loss in students, it is hoped that it will be possible to better understand, and hopefully mitigate, these longer-term impacts.

Research Matters / 34
A Cambridge University Press & Assessment publication

ISSN: 1755-6031

Journal homepage: https://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-matters/

Which assessment is harder? Some limits of statistical linking

Tom Benton and Joanna Williamson (Research Division)

To cite this article: Benton, T., & Williamson, J. (2022). Which assessment is harder? Some limits of statistical linking. *Research Matters: A Cambridge University Press & Assessment publication*, *34*, 26–41

To link this article: https://www.cambridgeassessment.org.uk/Images/research-matters-34-which-assessment-is-harder.pdf

Abstract:
Equating methods are designed to adjust between alternate versions of assessments targeting the same content at the same level, with the aim that scores from the different versions can be used interchangeably. The statistical processes used in equating have, however, been extended to statistically "link" assessments that differ, such as assessments of the same qualification type that assess different subjects. Despite careful debate on statistical linking in the literature, it can be tempting to apply equating methods and conclude that they have provided a definitive answer on whether a qualification is harder or easier than others.

This article offers a novel demonstration of some limits of statistical equating by exploring how accurately various equating methods were able to equate between identical assessments. To do this, we made use of pairs of live assessments that are "cover sheet" versions of each other, that is, identical assessments with different assessment codes. The results showed that equating errors with real-world impact (e.g., an increase of 5–10 per cent in the proportion of students achieving a grade A) occurred even where equating conditions were apparently favourable. No single method consistently produced more accurate results than the others.

The results emphasise the importance of considering multiple sources of information to make final grade boundary decisions. More broadly, the results are a reminder that if applied uncritically, equating methods can lead to incorrect conclusions about the relative difficulty of assessments.

Research Matters / 34
A Cambridge University Press & Assessment publication

ISSN: 1755-6031

Journal homepage: https://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-matters/

Progress in the first year at school

Chris Jellis (Cambridge CEM)

To cite this article: Jellis, C. (2022). Progress in the first year at school. *Research Matters: A Cambridge University Press & Assessment publication, 34*, 42–56

To link this article: https://www.cambridgeassessment.org.uk/Images/research-matters-34-progress-in-the-first-year-at-school.pdf

Abstract:
The results of an assessment taken at the start and end of the Reception Year by children in state schools in England over three years were analysed. Over 70 000 children were assessed during this time. The results of the analysis provided evidence of what the average child could do when they started school, and how much progress they made in that first year. Children typically start school with a wide range of skills and experiences and once they are settled into life in school, they make exceptional progress in their first year.

Research Matters / 34
A Cambridge University Press & Assessment publication

ISSN: 1755-6031

Journal homepage: https://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-matters/

What are “recovery curricula” and what do they include? A literature review

Martin Johnson (Research Division)

To cite this article: Johnson, M. (2022). What are “recovery curricula” and what do they include? A literature review. *Research Matters: A Cambridge University Press & Assessment publication, 34*, 57–75

To link this article: https://www.cambridgeassessment.org.uk/Images/research-matters-34-what-are-recovery-curricula-and-what-do-they-include.pdf

Abstract:

The concept of educational recovery is relevant to many systems, both those that experience some form of sudden disruption as well as those that historically have been prone to disruption. Our involvement in developing a curriculum framework for displaced learners in the Learning Passport project (UNICEF, 2020) made us more aware of the field of Education in Emergencies. An educational emergency is a situation where “man-made or natural disasters destroy, within a short period of time, the usual conditions of life, care and education facilities for children and therefore disrupt, deny, hinder, progress or delay the realisation of the right to education” (Committee on the Rights of the Child, 2008, p. 1). The COVID-19 pandemic has made the concept of emergency and recovery more relevant to even more education systems. The literature review described in this article was carried out to identify what recovery curricula are (e.g., what they seek to achieve, what information they cover, etc.), as well as to consider any evidence for their efficacy. By exploring the recovery curricula literature, we also wanted to consider the extent to which the concept is a singular, generalisable one, or whether it is tied to specific contexts.

Research Matters / 34
A Cambridge University Press & Assessment publication

ISSN: 1755-6031

Journal homepage: https://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-matters/

What's in a name? Are surnames derived from trades and occupations associated with lower GCSE scores?

Joanna Williamson and Tom Bramley (Research Division)

To cite this article: Williamson, J., & Bramley, T. (2022). What's in a name? Are surnames derived from trades and occupations associated with lower GCSE scores? *Research Matters: A Cambridge University Press & Assessment publication, 34*, 76–97

To link this article: https://www.cambridgeassessment.org.uk/Images/research-matters-34-whats-in-a-name.pdf

Abstract:
In England, there are persistent associations between measures of socio-economic advantage and educational outcomes. Research on the history of names, meanwhile, confirms that surnames in England – like many other countries – were highly socially stratified in their origins. These facts prompted us to wonder whether educational outcomes in England might show variation by surname origin, and specifically, whether surnames with an occupational origin might be associated with slightly lower average GCSE scores than surnames of other origins. Even though surnames do not measure an individual's socio-economic position, our hypothesis was that in aggregate, the educational outcomes of a group defined in this way might still reflect past social history.

In line with the research hypothesis, the results showed that the mean GCSE scores of candidates with occupational surnames were slightly lower than the mean GCSE scores of candidates with other surnames. The difference in attainment was a similar size to the difference expected between candidates half a year apart in age, and much smaller than the "gap" between male and female candidates. The explanation for the identified effect was beyond the scope of the current research, but surname effect mechanisms proposed in the literature include the psychological (e.g., implicit egotism), sociological and socio-genetic.