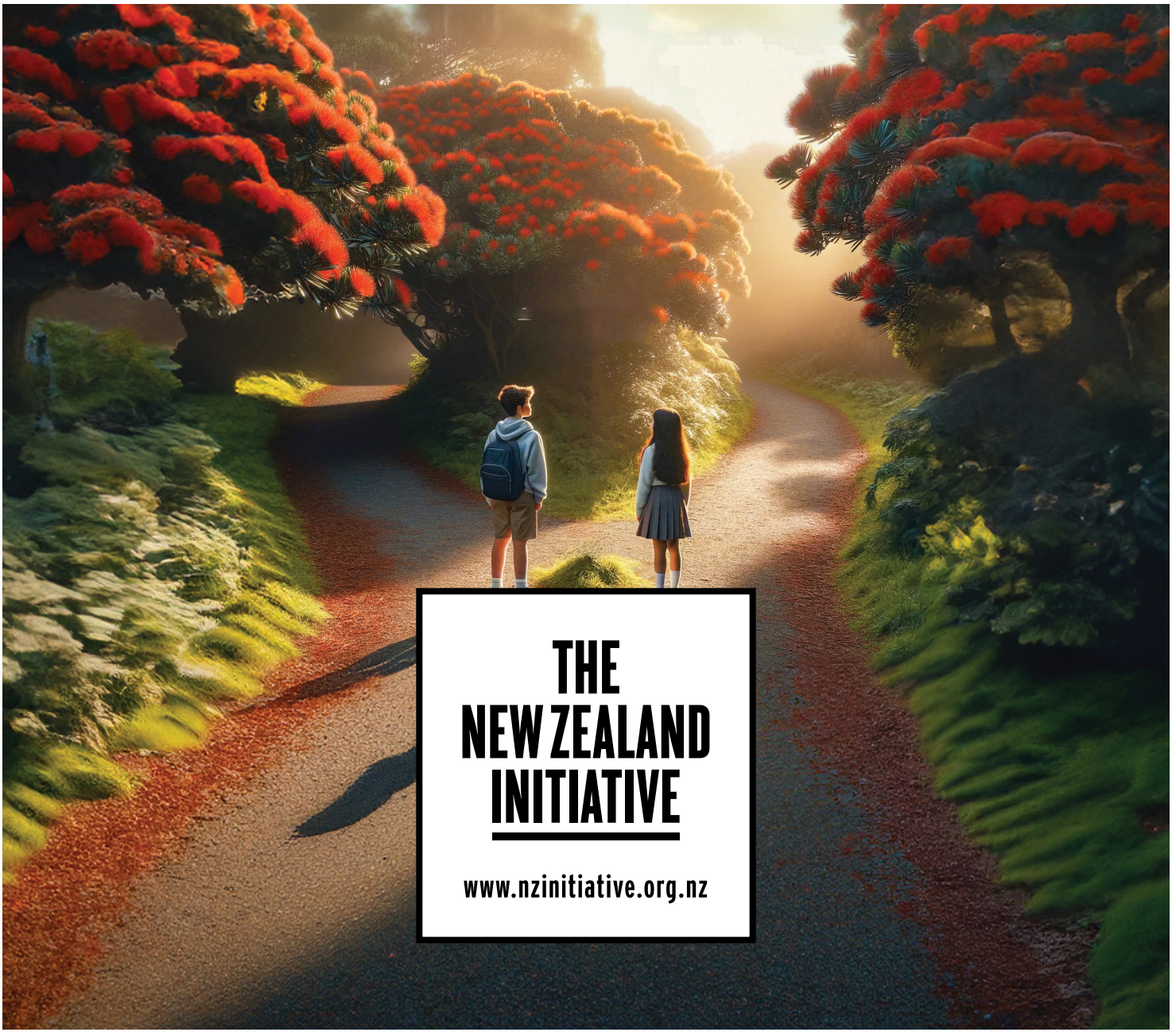
EDUCATION

CLASS DIVIDES?

The impact of streaming on educational achievement and equality

Michael Johnston
Benjamin Macintyre



THE NEW ZEALAND INITIATIVE

Published November 2023 by
The New Zealand Initiative
PO Box 10147
Wellington 6143
New Zealand
www.nzinitiative.org.nz

ISBN
978-1-7386009-8-4 (print)
978-1-7386009-9-1 (online)

RR79

Printing arranged by TBD Digital

CLASS DIVIDES?

The impact of streaming on educational achievement and equality

Michael Johnston
Benjamin Macintyre

About the New Zealand Initiative

The New Zealand Initiative is an independent public policy think tank supported by chief executives of New Zealand businesses. We believe in evidence-based policy and are committed to developing policies that work for all New Zealanders.

Our mission is to help build a better, stronger New Zealand. We are taking the initiative to promote a prosperous, free and fair society with a competitive, open and dynamic economy. We are developing and contributing bold ideas that will have a profound, positive and long-term impact.

ABOUT THE AUTHORS

Dr Michael Johnston
Senior Fellow

Dr Michael Johnston is a Senior Fellow at The New Zealand Initiative. He leads the workstream on education. Before joining the Initiative, Dr Johnston held academic positions at Victoria University of Wellington (2011–22), including Associate Dean (Academic) in the University's Faculty of Education (2020–22). He was a lecturer in psychology at the University Melbourne and a Research Fellow at Latrobe University. Dr Johnston holds a PhD in Cognitive Psychology from the University of Melbourne.

Benjamin Macintyre
Research Assistant

Benjamin Macintyre is a Research Assistant at The New Zealand Initiative. He will initially be focussing on education research. Benjamin has a BA (with honours) in Politics, International Studies and Italian and an MA (with merit) in International Security from the University of Warwick.

ACKNOWLEDGEMENTS

The authors are grateful for feedback and advice from Dr Harold Cuffe, Associate Professor Jack Robles, Professor John Hattie, Tim O'Connor, Dr John Etty and Stephanie Martin. These people made invaluable contributions to this report, including discussion, feedback and critical comments on initial drafts. They do not necessarily endorse the recommendations.

All remaining errors and omissions are the sole responsibility of the authors.

Contents

Foreword

Humans are set apart from other animals by our capacity for abstract reasoning, long-horizon planning and cooperation (hence *homo sapiens*, meaning "wise man"). Children are instinctively curious and have breathtaking potential, but cognitive limitations leave them dependent upon adults until they come of age. Thus, when we see children behave impulsively, we recognize that their internal development is what holds them back. When adults charged with their care act without sufficient evidence or deliberation, we have a moral obligation to insist on better.

In yet another clear-eyed report from the Initiative, researchers Michael Johnston and Benjamin Macintyre examine the discourse surrounding the recent push to remove streaming from New Zealand primary and secondary schools. Streaming, or the grouping of students with peers of similar academic attainment, is widespread in New Zealand schools, just as it is in many countries across the planet. There are many reasons to question the merits of streaming, in its many forms. These include its uncertain efficacy in improving academic outcomes, its potential to create inequities and possible negative consequences for relatively low achievers' self-esteem. Yet, streaming is prevalent precisely because there are also good arguments for its use.

This report's deep dive into the public debate on streaming in New Zealand's schools reveals, in fact, an absence of debate. Advocates for "de-streaming" have provided policymakers with largely anecdotal and qualitative evidence to support their position that streaming should be removed from New Zealand schools. However, the volume of credible research on streaming is shown to be equivocal, and apparently subject- and context-specific.

Despite the criticism directed at advocates of de-streaming, this paper is not a cheap reactionary "stop de-streaming" hit piece. (As shown in the report, de-streaming combined with after-school tutoring programs and improved curricula may serve everyone's needs better than the status quo.) Instead, the authors of this report rightly ask our education policy makers to use the tools of science to understand the plausible impact of removing streaming from schools before intervening in schools' decisions about streaming. The pairing of statistical hypothesis testing with randomized trials is a proven way to deduce cause and effect, yet there appears to be little inclination to systematically trial de-streaming before sweeping policy recommendations is made.

The guidance in this report is specific to streaming, but the basic wisdom is general: for good policy, define transparent objectives, consider multiple perspectives, collect more data and use proven quantitative methods before acting. This wisdom would be well headed in New Zealand, where we have a history of adopting education policies, often prematurely, with far-reaching negative or unintended consequences. (In the past, The Initiative has covered how a Ted Talk influenced classroom design in primary schools, the poor credibility of NCEA as a certification system, and our belated shift away from the failed Reading Recovery curriculum).

The moral justification for our authority over children derives mainly from the superior capacity of the adult human brain for patient and reasoned decision making. We should use that capacity more when it comes to education policy.

Dr Harold Cuffe
School of Economics and Finance
Victoria University

Executive Summary

Streaming is the practice of organising students into groupings based on prior attainment. It is longstanding and widespread in New Zealand schools. However, recent debates have highlighted research associating streaming with increased achievement gaps between students with higher and lower prior attainment. Some commentators have called for a ban on streaming. In this report we examine a range of New Zealand based and international research, to contribute to these debates.

As critics of streaming often point out, data showing greater gaps between higher- and lower-achieving students in streamed environments than unstreamed ones are commonplace in the research literature. However, the relationship between streaming and educational inequality is not necessarily direct or causal. Other factors might mediate this relationship, such as students in lower streams being taught from less rich curricula and by less experienced teachers. Nonetheless, an important consideration for streaming is its effects on the learning efficacy of students from groups stereotyped as having low academic capability.

Stereotype threat is the phenomenon that, when members of stereotyped groups are in situations that make those stereotypes salient, they can become self-fulfilling prophecies. In New Zealand, Māori, Pasifika and, in co-educational schools, boys, may all be vulnerable to stereotype threat. When students from these groups are placed in lower streams, negative stereotypes about their academic capability may be psychologically activated, leading to a decline in motivation and educational progress. There is a risk of a vicious cycle, with long-term impacts on students' academic trajectories and self-perceptions.

Schools that offer a rich, diverse curriculum, high-quality teaching, clear learning goals and plenty of formative feedback can mitigate some of the potential negative consequences of streaming, including stereotype threat. Research suggests that schools considering destreaming must also focus on these things. Destreaming efforts that fail to do so often do more harm than good.

Advocates contend that streaming optimises learning for all students. They argue that it is easier for teachers to select curriculum content and pace learning appropriately when the students in a class are all at similar points in their academic development. In streamed classrooms, high-achieving students can proceed at a faster pace, avoiding boredom or frustration, while those who require more support can progress at a pace that does not overwhelm or leave them behind.

Cognitive load theory provides a framework for teaching based on a scientific understanding of human memory and attention. Specifically, most human learning is mediated by *working memory*, which has a finite capacity and decays rapidly. When its capacity is exceeded, learning is impaired. This is known as cognitive overload. If cognitive overload is persistent, it can lead to confusion, frustration and demotivation. With sufficient practice and iteration, knowledge is encoded in long term memory, where it is no longer subject to the constraints of cognitive load.

Cognitive load theory, then, has important implications for teaching, especially in subjects with hierarchical epistemic structures, such as mathematics, early literacy and science. Streamed classrooms theoretically assist teachers to manage students' cognitive load effectively,

by enabling them to tailor curriculum content and the pace of instruction to a narrower range of student needs.

Stereotype threat and cognitive load theory provide contrasting accounts of the likely effects of streaming. Taken together, they show why debates about streaming are not straightforward. Schools that stream must find ways to mitigate the risk of stereotype threat and other potential effects on the learning efficacy of students placed in low streams. Those that do not stream must ensure that teachers are able to manage students' cognitive load in mixed attainment classrooms.

One of the key challenges for policymakers is differentiating the direct impact of streaming from the influence of associated factors, such as teaching quality, curriculum richness, and the socio-economic circumstances of students. Without a robust, methodologically sound research base, policy decisions may be based on incomplete and misleading information.

To date, there has been no large-scale, quantitative study of the prevalence, nature or effects of streaming in New Zealand. Existing New Zealand based research is primarily small-scale and qualitative. It rarely reports measured effects of streaming on students' learning, and authors frequently start from an assumption that streaming is an undesirable practice. We recommend a large-scale, empirical study to better inform the debate on streaming in New Zealand. Instead of forcing schools to abandon streaming, they must be informed by research that elucidates both the risks of streaming and the challenges of destreaming. That will better position schools to make informed decisions, taking into account their local needs and resources.

Recommendations

The Ministry of Education should commission research on streaming's prevalence, nature and educational effects in New Zealand. That research should then be published, and schools encouraged to adapt their practice according to the evidence it provides. That may entail cessation or modification of streaming practices.

CHAPTER 1
Introduction

Education is a cornerstone of modern society. A high-quality school system provides young people with knowledge and the means to pursue it for themselves. It prepares them for the world of work and has the potential to break inter-generational cycles of poverty. A sound education also prepares children with the knowledge and thinking skills they need to participate meaningfully in democratic society.

Universal and compulsory education was implemented in New Zealand in the late-19th century. The *Education Act 1877*[1] entitled all children to a free education and made it compulsory for non-Māori up to age 13. In 1894, it was made compulsory for Māori as well. Now, all New Zealanders must attend school between the ages of six and sixteen.

Since the inception of universal and compulsory education, schools in New Zealand and many other countries have commonly stratified their classes so that teaching can be focussed on children at relatively similar educational levels. In New Zealand, this practice is known as *streaming*. Streaming is broadly defined as the separation of students into classes based on their prior attainment.[2] Students with higher attainment are grouped together, as are their less advanced peers. Streaming is not a monolithic practice, however, and the term encompasses a broad range of educational practices. Various terms are used to describe the practice internationally. In the United States, it is often referred to as *tracking*. In the UK, it is sometimes known as *setting* or *ability grouping*.[3]

In New Zealand, decisions regarding whether or not to adopt streaming, and how it is implemented, are made at the level of individual schools. There is a lack of reliable evidence on streaming in New Zealand. The Ministry of Education holds no data on how widespread streaming is in New Zealand schools, nor on how schools that use streaming implement it. This makes it difficult to analyse the effects of different streaming practices on student achievement in New Zealand.

In contrast, the international literature on streaming is extensive and contentious, with debate on the practice ongoing since the early 20th century. Theorists' positions are sometimes influenced by their philosophical and political aims – especially on issues such as educational equality – as much as by evidence regarding its educational impact.[4]

The younger the children, the more controversial streaming becomes. R. Marks notes that teachers and parents are often opposed to designating children as having greater or lesser attainment when they have only just commenced formal education.[5] Opposition to formally streaming very young children is probably well-founded. Any teaching efficiency gains achieved through streaming must be weighed against potential for negative motivational and self-efficacy impacts on students placed in lower streams. The available information on the prior attainment of children who have just started school is very limited. It is also unreliable because accurate assessment of the capabilities of very young children is difficult. Furthermore, risks to students' self-beliefs may be especially acute early in their schooling, before they have established themselves as learners.

Various methods of streaming are documented in the literature. Students may be allocated to the same stream for all subjects or to different

streams for different subjects. The former approach is the most common internationally.[6] Most schools do not simply divide their student cohorts in two – usually, there are also intermediate streams. But how many streams there are, and how each is resourced and taught, is dependent on the size of a school, the number of teachers it employs, its educational culture and other factors.

A less formal approach is not to stream pupils into separate classes, but to place them in different instructional groups within the same classes.[7] For the purpose of our discussion, this approach is considered a type of streaming, because it involves students being grouped based on their previous attainment.

Internationally, streaming tends to be more common in secondary schools than in primary schools. For example, in the United States, 24% of primary schools use some form of streaming compared to 48% of middle schools and 40% of high schools.[8] This is probably true in New Zealand as well.

To understand the impact of streaming on students, it is important to understand how the attitudes and practice of teachers tend to differ in relation to students in different streams. A common argument against streaming is that schools tend to give students in higher streams more experienced teachers who provide more engaging pedagogy.[9]

Proponents of streaming argue that it maximises the educational progress of students. According to this argument, grouping students with similar levels of attainment enables all children to progress at rates that suit them: Students with greater prior attainment are not held back, and those facing academic challenges are not left behind. Streaming is efficient for teachers because it enables them to focus teaching more tightly on groups of students at similar stages of curriculum progress. In practice, however, the impacts of streaming are nuanced, multi-faceted and unpredictable. Its effects on academic achievement and students' learning efficacy vary and are subject to the specific contexts and ways in which it is implemented. Factors such as the psychological impact on students placed in low streams must be taken into account.

Streaming has recently been subject to much negative analysis and commentary in New Zealand. Several reports have been published that condemn the practice. Two such reports, *Ending Streaming in Aotearoa* and *Kōkirihia*,[10] have been especially influential, moving the national conversation towards banning streaming altogether. First-hand accounts of negative impacts of streaming on students – particularly those placed in lower streams – have led the debate. There has been a heavy focus on the impact of streaming on Māori and Pasifika students, who are over-represented in lower-stream classes.[11]

When he was Minister of Education, Chris Hipkins claimed that "streaming increases inequality"[12] and that it "does more harm than it does good".[13] The Ministry of Education would like schools to abandon streaming voluntarily. An internal Ministry of Education report outlined an aim to "to facilitate a national discussion about shifting towards more inclusive education practices and structures and away from streaming".[14] The Ministry's position is that streaming causes inequitable education outcomes and that its removal will improve student attainment.

The international academic literature on streaming is more equivocal in its findings than the tone of the New Zealand debate might imply. While many studies suggest that streaming often exacerbates educational inequality, others suggest that, in certain curriculum areas, especially mathematics and early reading, it can improve learning across the range of student attainment. Still others have shown that unless destreaming – the cessation of grouping students by prior

attainment – is accompanied by improved curricula, assessment and pedagogy, it can do more harm than good.

In this report, we distinguish between *streaming* and *tracking*. We define *streaming* as the separation of students into groups based on prior attainment, but nonetheless taught within the same broad curriculum structure. We define *tracking* as the practice of separating students into different cohorts and teaching each with different curricula, usually along academic and vocational lines.[15]

Some countries organise their school systems such that entire schools are allocated to one track or another.[16] For example, German secondary schools comprise four types: *Gymnasia* and *Gesamptschulen* have an academic focus and prepare students for university. *Realschulen* and *Hauptschulen* are geared towards preparing students for apprenticeships and vocations.[17] While between-school tracking is common in some countries, it is relatively rare in the anglosphere. In countries in which between-school tracking occurs, it is usually implemented at the systems-level, rather than school-by-school.

In New Zealand there is no formal system of tracking as there is in Germany. Even so, the NCEA system results in quasi-tracking. NCEA comprises two kinds of assessment standards: achievement standards, which assess academic subjects, including those that can contribute to University Entrance (UE)[18]; and unit standards, which tend to be vocational. Lower decile (higher Equity Index) schools disproportionately assess their students using unit standards. The differentiation of students into different curriculum paths according to their enrolment in NCEA standards is effectively a form of tracking.

Unfortunately the New Zealand streaming debate has included no focus on quasi-tracking through NCEA, which would be unaffected by a ban on streaming. Māori and Pasifika students are more likely than New Zealand European or Asian students to undertake non-academic (vocational or contextual) study for NCEA, especially at Levels 2 and 3.[19] This is reflected in data showing that Māori and Pasifika students complete far higher proportions of their NCEA assessment with unit standards than European or Asian students. This contributes to their much lower rates of attainment of UE.[20]

Choices in respect of NCEA options are ostensibly made by students and their parents. However, Jensen, Madjar and McKinley noted that students and parents are not always clear on the consequences of these choices for tertiary education opportunities.[21] University Entrance, for example, depends on completing sufficient credits in approved subjects. These authors commented that "Māori, Pacific and lower decile secondary school students are particularly at risk of ending up with inappropriate choices [of NCEA courses] if … they do not receive clear guidance about the best course choices".[22] Schools, then, need not formally track students for curriculum differentiation to occur – and this differentiation is correlated with students' ethnicities and socioeconomic situations.

This report examines both New Zealand and international research and debate to clarify different ways streaming is implemented and the impacts of these different approaches. The literature review and discussion we present could inform the design of a large-scale, quantitative study into the prevalence and effects of streaming. Such a study, which has never been carried out in New Zealand, would be of great assistance to policymakers and schools. In the light of the evidence it would provide, some schools may decide to end streaming. For those that decide to continue the practice, it would alert them to some of its risks and potentially inform them of modifications that could enable them to better meet the learning needs of their students.

CHAPTER 2

The streaming debate in New Zealand

Most New Zealand secondary schools stream,[23] but there is no public repository of information on which schools, how streaming is typically implemented, for which subjects, or at what age levels. Neither is there any information on whether students in lower streams get more or fewer resources, or the extent to which students move between streams. Most critically, there has been no systems-level analysis of the causal impact of streaming on the educational progress of New Zealand students.

To demonstrate a causal relationship – either positive or negative – between streaming and educational outcomes, it must be shown that students who are streamed have, on average, different outcomes than those who are not streamed when other factors are held constant. This requires a large-scale quantitative research design. To date, no such study has been conducted in New Zealand. Most New Zealand-based research on streaming has used qualitative designs.

The most robust approach to examining the educational impact of streaming would be to use a true experimental design. An experimental approach to understanding the educational impact of streaming would involve students being allocated to streamed and non-streamed educational environments at random, and quantitative measurement of subsequent educational progress. If, as most New Zealand commentators claim, streaming is causally implicated in poor educational outcomes, students allocated to unstreamed environments should make better progress than those allocated to streamed environments.

Provided a sufficiently large sample is employed, the random allocation process ensures that all factors other than the one of interest – whether or not students are streamed – are approximately evenly distributed within each of the streamed and unstreamed groups. Any observed difference in outcomes could therefore be attributed to streaming. Outcomes for different streams could be measured separately, to determine whether any detrimental effects of the practice differentially affect students in lower streams. Similarly, outcomes can be compared across streamed and unstreamed conditions for sub-groups of students, such as Māori.

Unfortunately, true experiments of this nature are difficult to run in New Zealand's educational environment. Schools are self-governing and cannot be compelled to participate in research. Many might regard implementing experimental initiatives as unfeasibly disruptive to their operations. Nonetheless, known correlates of educational achievement – such as socioeconomic variables – can be statistically equated in streamed and unstreamed samples. This, however, leaves unknown correlates as potential explanations for observed differences in outcomes.

Whether extraneous variables are controlled experimentally or statistically, research seeking to determine the effects of streaming must employ large samples and quantitative measures of educational outcomes. Qualitative research – currently a majority of New Zealand-based streaming research – typically employs small samples and, by definition, does not involve quantitative measures. It is therefore unsuitable to identify causal relationships between streaming and educational attainment.

Qualitative evidence has additional limitations for policy decisions.

> Qualitative research is frequently criticised for lacking scientific rigour with poor justification of the methods adopted, lack of transparency in the analytical procedures and the findings being merely a collection of personal opinions subject to researcher bias.
>
> — Noble and Smith[24]

As Noble and Smith note, there are methods that can mitigate these weaknesses: accounting for and acknowledging certain biases in the research; establishing frameworks for comparison with other perspectives; and triangulating data with other data sets to verify the validity of findings.

There are two fundamental limitations of qualitative research for its capacity to support causal claims. One relates to sampling procedures: Qualitative samples are rarely representative of the populations of interest and their typically small size makes them unreliable. The other relates to measurement: Qualitative studies do not measure educational outcomes or progress. By its nature, qualitative research relies on the self-report of participants. Even if the self-report of streamed and unstreamed students were to be compared – which is not typical in qualitative designs – there would be no way to rule out extraneous variables as explanations of any differences between streamed and unstreamed students. Qualitative research cannot, even in principle, contribute to establishing any causal link between streaming and educational outcomes.

This is not to say that qualitative research is worthless. It can act as a useful starting point for quantitative work by identifying issues to follow up using rigorous measures. It can also be a valuable supplement to quantitative work by more deeply probing potential reasons for any observed differences in outcomes. These approaches are known as 'mixed methods' research designs because they employ both qualitative and quantitative procedures. Mixed methods designs combine the strengths of quantitative and qualitative methodologies. Policy decisions should not, however, rest on qualitative data alone.

In the rest of this chapter, we review some of the most influential recent commentary and research on streaming in New Zealand. This provides both insight into current discourse on the issue and useful background for the large-scale quantitative research that would be required prior to any policy decision regarding banning streaming in New Zealand schools.

Tokona Te Raki

The Māori advocacy group, Tokona Te Raki, has been active in its efforts to ban streaming, and has published two recent reports on the topic, *Ending Streaming in Aotearo*a and *Kōkirihia*.[25, 26] These reports have had a substantial influence on the Ministry's policy direction, the latter report having been partially funded by the Ministry.

Tokona Te Raki seeks to improve outcomes for Māori across three different sectors, describing themselves as using "social innovation to achieve equity in education, employment and income for Māori".[27] They seek to "create long-term structural solutions so that change is both scalable and sustainable".[28] They are not attached to any university, and their reports are written to be accessible to the general population.

Māori are over-represented in statistics related to poor educational outcomes, and Māori and Pasifika students are over-represented in lower streams.[29] While the educational discourse in New Zealand has tended to focus on cultural differences[30] and racism[31] as sources of Māori educational disadvantage, socioeconomic differences are likely to be an important part of the problem. Socioeconomic gradients in education are well-established internationally, with the children of parents with low occupational status typically having poorer outcomes than those of parents with higher-status employment.[32]

Average Māori incomes are lower than average non-Māori incomes[33] and in 2022 the Māori unemployment rate was 1.9 times that of the rate for non-Māori.[34]

Whatever its causes, Tokona Te Raki seeks to rectify persistent inequalities in educational success for Māori. The conclusion of both *Ending Streaming in Aotearoa* and *Kōkirihia* is that streaming is an inequitable educational practice that exacerbates poor outcomes for Māori students. However, Tokona Te Raki's original work on streaming is overly reliant on qualitative evidence. There is some attempt in *Ending Streaming in Aotearoa* to use quantitative data, particularly in respect of their findings about the ending of streaming in Horowhenua College. Data from that school showed that results in mathematics improved for Māori and Pasifika students after streaming was discontinued. However, at the same time they ended streaming, Horowhenua College changed its methods of assessment and established mentoring and tutoring programmes. It is possible that it was these measures, and not destreaming, that resulted in the improved outcomes.

In *Kōkirihia*, the evidence is entirely qualitative. Interviews with teachers and students with a negative view of streaming, or who felt negatively affected by streaming themselves, form the backbone of *Kōkirihia*'s evidence base. Some of the evidence from teachers is self-described as anecdotal.[35] The authors did not acknowledge or account for either their own biases or those of their respondents. Neither was there any apparent attempt to canvass countervailing views or to acknowledge the limitations of their research methodology.

In their literature review, the authors of *Kōkirihia* claim that "research evidence is *unequivocal* that fixed ability grouping in any form does not work for the vast majority, and any advantages for high achievers are minimal".[36] In support of this claim, they cite eight studies.[37] While this body of research is clearly not in favour of streaming, the Tokona Te Raki authors' claim that it presents "unequivocal" evidence against streaming is overstated. Here, we examine the research evidence from the sources cited by Tokona Te Raki to draw out some nuance.

Elley[38] reviewed a range of early-to-mid 20th century quantitative, quasi-experimental studies on streaming. A preponderance of this evidence suggested little overall effect of streaming on educational attainment. Where attainment was shown to be affected, it tended to improve outcomes for students in higher streams, but at the cost of poorer attainment for lower streams. This accords with the broad findings of Gamoran's review,[39] and with Hattie's meta-analysis.[40] The latter showed that the net effect of streaming was close to null, but Hattie noted that this masks a tendency to increase learning gaps between students allocated to higher and lower streams. Hanushek and Woessmann,[41] also found that early streaming tends to increase inequality in achievement.

Not all the research cited in *Kōkirihia* supports the hypothesis that streaming exacerbates educational inequality. Ireson et al. conducted a study of 6,000 students studying for the General Certificate of Secondary Education (GCSE) in England and Wales, in English, mathematics and science.[42] This study showed almost no effect of streaming based on the number of years GCSE students had been in streamed classes, irrespective of their level of prior attainment. There was, however, a very small tendency for higher-attaining students to attain *lower* grades in GCSE science the longer they had been in streamed classes, whereas lower-attaining students attained slightly *higher* grades following more years in streamed classes. This suggests that, in the context of GCSE science, streaming slightly *reduced* educational inequality. This is the reverse of the general pattern in the research reviewed by Elley[43] and Gamoran[44], and in Hattie's meta-analysis[45]. It does, however, agree with some other subject-specific analyses.

Gamoran cited studies by Slavin[46] and Conner and colleagues[47] suggesting that, in mathematics and early reading respectively, differentiated instruction can be beneficial for students with lower prior attainment.

Another common finding in Elley's review was that the social development of non-streamed students tends to be better than that of their streamed peers. Even so, while Elley concluded that streaming has "not really proven itself",[48] he did not conclude that it is unequivocally harmful. Rather, he noted a need for properly controlled studies before drawing firm conclusions. Nearly 50 years later, that need remains unfulfilled in the New Zealand context.

Gamoran argued that the propensity of streaming to amplify educational inequality interacts with a general socioeconomic gradient in educational attainment: Because students from lower socioeconomic backgrounds tend to make slower educational progress than their more affluent peers, they are more likely to be placed in lower streams.[49] If lower streams do not receive appropriate and targeted pedagogy, streaming can therefore exacerbate pre-existing educational disadvantage based on socioeconomic factors. And because many ethnic minorities are over-represented in lower socioeconomic strata, any negative impact of streaming falls disproportionately on students from those minorities.

Glock et al. found that teachers tend to be less accurate and less confident in making school placement judgements for ethnic minority students than for ethnic majority students.[50] Such stereotypical expectations are likely to result in misallocation of ethnic minority students to lower-streamed classes and, thereby, to exacerbate any inequitable effects of streaming for those minorities.

Elley, Gamoran and Hattie all emphasised the importance of effective classroom instruction, irrespective of streaming.[51] Gamoran, in particular, noted that lower-streamed students tend to receive less stimulating teaching from less experienced teachers. He cited research from Taiwan,[52] Israel,[53] and Scotland,[54] showing that educational inequality can be reduced in streamed environments when lower-streamed students are provided with clear standards and regular and meaningful assessment. Elley also noted that detrimental effects of outcomes for lower-streamed students are often attributable to their being allocated to less experienced teachers.[55] Hattie noted qualitative research suggesting that "…low track classrooms are more fragmented, less engaging, and taught by fewer well-trained teachers". He further argued "that the quality of teaching and the nature of the student interactions are the key issues, rather than the compositional structure of the classes".[56]

Cessation of streaming entails challenges. Teaching classes comprising students at different curriculum levels can be technically difficult. Gamoran[57] noted that failure to adequately address that difficulty can result in lower-achieving students being left behind, and a generally less interesting and challenging curriculum. Any policy decision to discontinue streaming would need to be accompanied by preparation of teachers to provide meaningful and well-targeted instruction in mixed attainment classrooms. Initial teacher education in New Zealand does not typically provide the pedagogical content knowledge required for this preparation.[58]

The most salient pattern in the research evidence cited in *Kōkirihia* is that streaming does little for overall achievement and exacerbates educational inequality. Furthermore, because students from many ethnic minorities – in New Zealand, Māori and Pasifika – are more likely than other students to be placed in lower streams, they are more likely to suffer ill effects from streaming. The evidence is not, however, unequivocal; nor is it entirely representative of wider literature on streaming – of studies using experimental designs in particular.

Several studies cited in *Kōkirihia* note that it is likely to be practices commonly associated with streaming, rather than streaming itself, that is behind the disadvantage for low-streamed students. For example, we noted Elley's finding, nearly 50 years ago, that low-streamed students were typically given less stimulating pedagogy and less experienced teachers than their higher-stream peers.[59] There is no current systematic evidence in the New Zealand context to tell us whether this is still so. But if poor curriculum decisions or poorly designed pedagogy drive the inequitable outcomes often observed in the research, an alternative to banning streaming would be to adopt more stimulating curricula and more appropriate pedagogy. Slavin's finding that grouping students by attainment level can lift the achievement of lower-achieving students in mathematics if it is accompanied by well-designed pedagogy,[60] and O'Connor and colleagues' similar finding for early reading, lend credence to this possibility.[61]

In light of the nuances we have discussed here, it seems premature to call for a general ban on streaming. Indeed, doing so without addressing pedagogical issues – in particular, training teachers to use effective pedagogy and to cope with mixed-attainment classes – might do more harm than good.

Although two of the studies cited in cited in *Kōkirihia* – Elley and Hattie[62] – were by New Zealand-based researchers, none of the literature cited in the report was specific to the New Zealand context. We turn now to research examining the effects of streaming in New Zealand.

Research on streaming in the New Zealand context

To the best of our knowledge, no large-scale quantitative study of the effects of streaming has been conducted in New Zealand. The research that has been conducted tends to be qualitative, to employ small samples, and, often, to focus on attitudes and beliefs about streaming, rather than directly on its educational effects. A full review of this body of literature is not undertaken in this report. Instead, we focus on just three typical studies, to illustrate the limitations of the New Zealand research literature for informing policy decisions on streaming. These studies are all cited in a 2021 literature review from the Ministry of Education[63] and feature prominently in the bibliographies of other New Zealand streaming research.

Anthony and Hunter surveyed two cohorts of mathematics support teachers – 102 in total – to probe their beliefs about streaming and the extent to which they use it.[64] Findings indicated that streaming was a dominant method of organising classes in the schools surveyed, although not the only way. Most respondents opposed fixed grouping, with just 22% of Year 1–5 teachers and 16% of Years 5–8 in support. Even so, only a minority of respondents – ranging from 25% to 33% across the individual schools using grouping – disagreed with any attainment-related grouping in mathematics at all. The authors attributed the high level of support for mixed and flexible grouping to teachers' lack of knowledge of alternatives. This suggests that they had taken an *a priori* position against streaming, rather than taking seriously that teachers might have sound reasons for supporting flexible grouping.

Turner et al. explored links between teacher expectations and student ethnicity.[65] They surveyed fifteen mathematics teachers and 361 students from Auckland schools. Teachers displayed different expectations for students of different ethnicities, with expectations highest for Pakeha and Asian students and lowest for Māori students. The authors argued that, due to these lower expectations, Māori are more likely than students of other ethnicities to be inappropriately relegated to lower streams. This finding agrees with that of the international research of Glock et al.,[66] cited in *Kōkirihia*.

While this research adds to the evidence that any inequitable effects of streaming are likely to fall disproportionately on Māori students, it does not of itself comprise evidence that streaming is an inequitable practice.

Hornby and Witte surveyed fifteen Christchurch high schools ranging in roll size, student demographics and decile.[67] Like the study of Anthony and Hunter,[68] this study comprised a small sample, but surveyed schools rather than teachers. Each school in the sample was asked five questions: whether the school had a written policy on student grouping; what types of grouping were in use; what the perceived benefits of the schools' class arrangements were; what the perceived disadvantages were; and how grouping arrangements impacted learning for specific groups of students. The findings were mixed. Ten schools did not have a written policy while five did. Only one school did not group students based on prior attainment. In the fourteen other schools, approaches included special needs classes, high-, middle- and low-attainment classes, and mixed-attainment classes. One school had a high-attainment class for each year group and grouped everyone else together. This research speaks to the prevalence of streaming, although the small sample size restricts the reliability of those data. Data on the educational effects of streaming were limited to the qualitative report of respondents.

New Zealand commentators on streaming often cite studies as supporting claims in ways that are unwarranted by their evidence. Anthony and Hunter's study, for example, is widely cited in support of arguments against streaming.[69] Yet, it reports only a survey of what teaching assistants *think* of streaming, as opposed to its measured effects.

An assumption that streaming is harmful, especially for Māori and Pasifika, is nearly ubiquitous in the New Zealand literature. The international evidence, especially that of Glock et al.,[70] discussed above, lends credibility to that assumption. However, there is no generalisable New Zealand-based evidence that clearly evinces it, notwithstanding small-scale qualitative studies like that of Turner et al.[71] New Zealand researchers frequently cite earlier studies making the same assumption. The assumption itself is never seriously questioned.

To better inform the streaming debate in New Zealand, a large-scale, quantitative study is required. Such a study should focus on measuring the effect of streaming on achievement in different subjects, at different levels of achievement, for different student demographics, and at different points in schooling.

CHAPTER 3

International research on the prevalence and equitability of streaming

In the absence of any large-scale quantitative studies on the effectiveness of streaming in New Zealand, we must turn to the international research literature. In Chapter 2 we reviewed some international studies cited in Tokona Te Raki's report, *Ending Streaming in Aotearoa*.[72] As we noted, a preponderance of the evidence in those studies suggests that streaming has a negative effect on educational equality. Students allocated to higher streams tend to do better than their unstreamed peers. Those allocated to lower streams tend to do less well than theirs.

In this chapter, we consider further international quantitative research on streaming to inform the current debate in New Zealand. We identify some issues that should be followed up by research in the New Zealand context before any move is made to ban streaming. Some studies focus on the social and psychological effects of streaming in addition to, or rather than, its impact on learning. In this report, we limit the scope to the effects of streaming on learning itself.

We first review evidence from the Programme for International Student Assessment (PISA) run by the Organisation for Economic Co-operation and Development (OECD) to compare the prevalence of streaming in New Zealand with its prevalence internationally. We also use PISA research as a preliminary examination of the relationship between streaming, socioeconomic disadvantage and educational equality across the education systems of participating countries. We then compare the findings of several influential meta-analyses of quantitative research on streaming. We review individual studies to shed light on the effects of streaming in specific contexts, in particular, the evidence that streaming has different impacts in different subjects. Finally, we review some of the evidence on factors governing successful destreaming.

Evidence from PISA

The OECD's PISA data provides the widest reliable source of quantitative evidence available for comparison of education systems internationally. Every three years since 2000 (except during the COVID pandemic), PISA assessments have been run to measure the capabilities of 15-year-olds in mathematics and reading. Science was added in 2006. In 2018, some 600 thousand students from 79 countries participated in the programme.

As well as testing knowledge in these three curriculum areas, PISA participants complete questionnaires to measure their attitudes, backgrounds and experiences. PISA also collects data on school demographics, school types, and other metrics. Most importantly for our purposes, the OECD collects data on the prevalence and nature of streaming in the education systems of participating countries (the OECD refers to it as *ability grouping*).

According to a survey of principals run as part of the PISA cycle, in 2012, 67% of students were streamed in at least one subject and 46% were streamed in all subjects.[73] There is wide international variation in the prevalence of streaming, with New Zealand at the upper end of the distribution.

Many schools, both in New Zealand and internationally, stream for mathematics even if they stream no other subject. In some cases, students in different streams study similar content, and in others, different content is taught to different streams. New Zealand was one of ten countries in which more than 95% of secondary students were streamed for mathematics,[74] and in 2018, 83.5% of New Zealand secondary students attended schools that streamed some or all subjects.[75]

Socioeconomically disadvantaged students are more likely to be placed in less 'academic' streams than more advantaged students. This is unsurprising given well-documented socioeconomic gradients in educational achievement. The OECD notes that there is potential for an associated stigma or inadequate allocation of resources to lower-streamed students, which would thereby impact most heavily on the achievement of less advantaged students. Given that low-socioeconomic students are more likely to be assigned to lower streams – based on their lower average prior attainment – they are also at greater risk of any deleterious effects of streaming. The risk, then, is of streaming perpetuating and exacerbating socioeconomically mediated educational inequality.

Socioeconomically disadvantaged students are more 'resistant' to the impact of their circumstances on educational achievement in some countries than in others, meaning that the socioeconomic gradient in those countries is less steep. Greater resistance is typically observed in systems with less streaming.[76] Vandenberghe noted, though, that, in respect of mathematics learning, advantages to low-attaining students in unstreamed environments can come at a cost to the progress of higher-attainment students.[77] Benito et al. concluded that, while educational inequality is exacerbated by streaming, the effects on average attainment depend partly on school curricula, teaching, and approaches to grade (year-level) advancement.[78]

Following Mons' classification of schooling models,[79] Benito and colleagues' analysis showed that schooling systems using what Mons called a comprehensive *à-la-carte integration model* show relatively little impact of streaming on aggregate attainment.[80] However, gains attributable to streaming for higher-attainment students tend to be offset by losses for lower-attainment students. Hallmarks of this model are a common core curriculum until the senior secondary years, with automatic progression from one year-level to the next, and a lack of streaming at primary school level. New Zealand was classified in this grouping.

On the other hand, countries classified under Mons' *uniform integration model* tended to show higher overall attainment in unstreamed environments. The main difference between the à-la-carte and uniform models is that under the former, students are promoted from one grade (year-level) to the next irrespective of attainment, whereas under the latter, students remain in a grade if they have made insufficient progress.

In other analyses of PISA data, Agasisti and colleagues found that streaming can negatively impact socioeconomically disadvantaged students.[81] The age at which students were streamed modulated the extent of that impact; the earlier children were streamed, the poorer the outcomes for those students tended to be.

The utility of PISA data for elucidating the effects of streaming is limited. There are many variables that contribute to variation in PISA results across participating nations, and attempting to disentangle the impact of any one variable is statistically fraught. The researchers cited here all used statistical techniques to account for variables other than streaming that may mediate the correlation between streaming prevalence and educational inequality. However, these techniques account only for variables entered in the statistical models and leave open the possibility that other, unknown variables are

also involved. The only way to reliably control for all extraneous variables is to use a true experimental approach.

Meta-analyses

Meta-analyses of streaming research involve synthesising data across multiple studies to determine its average effects. The combined sample sizes of all the studies included in a meta-analysis affords much greater reliability than is typical of individual studies. On the other hand, meta-analyses tend to average out contextual effects in the individual studies, which can mask important nuances. If for example, streaming is effective in some subjects and harmful in others, and if a meta-analysis included approximately equal numbers of studies focussing on subjects in each of those categories, the net effect would likely be near zero. Unless specifically probed, differential effects of streaming across subjects would often be washed out by averaging across them.

Overall effect sizes of streaming in meta-analyses are often close to zero. This was the case in Hattie's meta-analyses,[82] Slavin's synthesis,[83] and, for between-class streaming at least, the second-order meta-analysis of Steenbergen-Hu et al.[84] Rui's meta-analysis of fifteen studies run between 1972 and 2006 showed better achievement for unstreamed students than for streamed students, although the effect size was modest.[85]

While these meta-analyses all showed little or no impact of streaming on overall achievement, they differed in their findings regarding the impact of streaming for high- and low-attainment groups taken separately. Hattie noted that streaming worsens educational inequality.[86] However, in Slavin's synthesis of 29 studies comparing students in streamed and unstreamed classes, effects for all attainment groups were statistically indistinguishable from zero.[87] In fact, there was no significant difference between the academic outcomes for streamed and unstreamed students in any of the individual studies included in his analysis.

Slavin's findings contrasted with those of Rui's study, which included sub-analyses of outcomes for students at different levels of attainment, and further sub-analyses comparing effects for different attainment levels using experimental studies only.[88] Using the full set of studies, Rui found little or no apparent impact of streaming on high- or average-achieving students.[89] Low-achieving students performed better in unstreamed classrooms, although the effect size was, again, modest. Only four experimental studies were included in Rui's analysis. A sub-analysis of these studies showed weak evidence that high- and average-attainment students achieved more in streamed than unstreamed environments, and stronger evidence that low-attainment students achieved substantially better in unstreamed environments.

The second-order meta-analysis of Steenbergen-Hu et al. covered over a century of research into streaming.[90] This analysis evinced no effect of streaming on overall attainment, nor any effect on the gaps between high and low-attainment students, when students of different attainment are allocated to different classes. However, students grouped *within* classes benefitted relative to ungrouped peers, irrespective of their attainment level. This finding suggests that any tendency for streaming to increase educational inequality might be eliminated if students in the different groups all have the same teacher. Alternatively, lower-attainment students might benefit from the presence of higher-achieving peers in their classes.

The meta-analyses reviewed here show mixed evidence regarding the relationship between streaming and educational equality. In a review of research literature on streaming, Ireson and Hallam concluded that, "The conflicting results of … meta-analyses suggest that ability grouping impacts on or operates in concert with other

factors to produce its effects".[91] This observation emphasises the importance of considering contextual factors that might mediate any causal relationship between streaming and the range of educational attainment.

Individual studies

Many individual quantitative studies have suggested that streaming can exacerbate differences between higher- and lower-attaining students. Like many meta-analyses, a common finding is that higher-attainment students either do better in streamed environments than unstreamed ones or are unaffected by streaming, and that lower-attainment students do less well in streamed environments.

Parsons and Hallam examined a sample of 19,000 children in the UK, taken from the Millennium Cohort Study.[92] They focussed on the results of these students on Key Stage 1 Assessment tests, which measure reading, writing, maths and science attainment. They found that, in schools that stream students, those allocated to high-attainment groups made significantly more progress than students with similar attainment at non-streaming schools. Their middle- and low-attainment peers achieved significantly less.

In a similar vein, Trinidad and King analysed the Philippines PISA data.[93] They found that streaming in the Philippines, which is widespread, did not produce better average academic outcomes than non-streaming schools, and was associated with wider disparities in academic achievement. In interpreting these results it is important to note that the PISA data do not enable researchers to determine which students were in which attainment groups. As Trinidad and King pointed out, this limitation made it impossible to identify the locus of the greater inequality in streamed environments – whether it was because high-attainment students did better, because low-attainment students did more poorly, or a combination of both.[94]

Slower-than-expected progress for low- and, often, intermediate-streamed classes, such as that in the studies cited above, is often used to argue against streaming. However, Hong, Corter, Hong and Pelletier noted that students in higher streams have learned faster than students in lower streams in the past – which is why they were placed in the higher streams.[95] Therefore, lower-streamed students making slower progress than higher-streamed students given the same instruction cannot necessarily be attributed to streaming. To demonstrate deleterious effects of streaming, it is not enough to show that streamed students in low-attainment groups make less progress than those in high-attainment groups, nor to compare their progress with normative progress. It must be shown that streamed students in low-attainment groups make less progress than those of comparable prior attainment in unstreamed environments. Some studies suggesting a negative impact of streaming on educational equality meet this requirement.

In a study of mathematics attainment involving 1,730 seventh-grade students in twelve Israeli high schools, Linchevski and Kutscher found no significant difference between the average scores of high-attainment students in streamed and unstreamed classes.[96] However, the scores of students of intermediate- and low-attainment students were significantly lower in streamed classes. Kerckhoff compared British students' achievement in reading and mathematics between streaming and non-streaming schools.[97] High-attainment students in streamed schools made greater progress over a five-year period than those of comparable prior attainment in unstreamed schools. Low-attainment students in streamed schools made less progress than their counterparts in unstreamed schools.

Not all studies have shown negative effects of streaming on educational equality.

For example, Cheung and Rudowicz collected the data of 2,720 junior high school students in Hong Kong.[98] The authors hypothesised that placing students in lower-attainment groups negatively impacts their self-esteem. They further hypothesised that this impact causes them to perform less well academically than they would have, had they not been streamed. They hypothesised converse outcomes for students in high-attainment streams. However, they found that both low- and high-attainment groups made slightly more progress in streamed classrooms than in unstreamed ones. They also found that streaming had little or no impact on students' self-esteem.

Contextual factors in the implementation of streaming

As we have noted, meta-analyses average the results of the studies they include, which often cancels out positive and negative contextual effects across those studies. We now turn to studies that shed light on some of these contextual factors.

One important consideration is that streaming might have different effects in different subjects. In New Zealand and elsewhere, primary schools often group students within the same class for reading and mathematics, but not for other subjects. There is some evidence to support this approach. In a literature review focussing on elementary (primary) schooling, Slavin found that grouping within classes is most effective when it is done for at most two subjects – typically mathematics and reading.[99]

Some experimental and quasi-experimental evidence also favours within-class attainment grouping for primary school mathematics. For example, Borg compared mathematics learning in two geographically and demographically similar elementary schools in Utah, one using within-class grouping, and the other, not. He found reduced educational inequality in grouped environments.[100] Low- and average-attainment groups made more progress than ungrouped peers, whereas high-attainment groups made similar progress whether they were grouped or not. Slavin found that, at primary-school level, attainment grouping within classes, rather than between-class streaming, raises achievement and does not exacerbate educational inequality.[101] He also emphasised the importance of regular reappraisal of groupings and the importance of differentiating curriculum and pedagogy to address the needs of the different groups.

In another review of research literature, in this case on streaming in secondary schools, Slavin concluded that there is little evidence that streaming mathematics incraeses the range of attainment in the subject.[102] He noted, however, that the evidence on this only applies to students in different streams being taught the same curriculum material; the experimental studies he reviewed did not compare high- and low-attainment streams being taught different material. Forgasz notes the potential for the latter approach to limit the ability of lower streams to advance in the subject.[103] This point echoes the finding of Gamoran; that apparent effects of mathematics streaming on educational equality can be partly explained by students in higher streams studying more mathematics and being taught more advanced material.[104]

Within-class grouping for reading has also been shown to be effective when implemented using carefully designed instructional approaches. Using an experimental design in Florida elementary schools, Connor et al. randomly allocated classes to a condition in which children were taught in small groups, differentiated by initial vocabulary measures.[105] These students, especially those with low initial vocabulary scores, made significantly more progress in reading than those in a control group that was not differentiated. A critical component of the success of this approach was carefully differentiated pedagogy, appropriate to the attainment level of each group.

The work of other researchers[106] also emphasises the importance of targeted pedagogy to avoid deleterious effects of attainment grouping on educational equality.

In other subjects, especially the humanities at secondary level, streaming seems less effective and more likely to impact negatively on educational equality. For example, in an experimental study of the effects of streaming on learning in 11th grade American history, Thompson found a large effect favouring mixed-attainment classes.[107]

Another contextual effect on streaming is the types of students who are differentiated. When most of the students in a school are not streamed, there is evidence that creating a 'gifted class' can greatly improve the academic progress of the students in that class without harming the progress of others.[108] Special classes for students with learning delays or disabilities can also help those students. For example, Ballis and Heath examined the impact of a Texan policy requiring schools to reduce enrolment into Special Education (SE) classes, so that no more than 8.5% of students in each school were in these classes.[109] They found that the removal of students from SE classrooms reduced the odds of those students attaining entrance to tertiary education by 37%. Moreover, they found a correlation between this policy and poorer overall performance for students in mainstream classes. This is likely attributable to the increased workload for teachers associated with catering to the students with learning disabilities, detracting from the time available to spend with other students. Finally, Ballis and Heath found that removal of children from SE classes disproportionately affected minority students.[110] They also hypothesised that parents of minority students with special needs might be less likely to have the resources to challenge the removal of their children from SE classes.

'Gifted' and 'Special Education' classes are not typical approaches to streaming in New Zealand (or elsewhere). Even so, while the studies of Card and Giuliano[111] and Ballis and Heath[112] shed little light on the effects of more general stratification, they suggest that the typical New Zealand approach of integrating both gifted and learning-disabled students in mainstream classes may not be in the best interests of those students, and in the latter case, of other students either.

Destreaming

Destreaming refers to moving from stratified to mixed-attainment classes. It is often implemented in the name of improving educational equality. Here we review three studies, two reporting improvements in academic achievement after stopping streaming,[113] and another observing deterioration.[114]

Burris and Wellner examined a multi-year effort to destream in the Rockville Centre School District on Long Island (near New York) in the late 1990s.[115] Homogenous grouping practices were eliminated, and all students were placed in mixed-attainment classes. Results were positive. In the three years before the change, 23% of African American and Hispanic students had passed a benchmark algebra examination. After the change, 75% passed. For White and Asian-American students, the rise was from 54% to 98%. While wider New York averages were also registered, the improvements for the Rockville students were much greater.

Several other interventions accompanied the cessation of streaming. Perhaps the most important was that the curriculum previously reserved for the high-attainment stream was adopted for all students, although Burris and Wellner noted that the mathematics curriculum was "revised and condensed".[116] Also, additional mathematics workshop classes and after-school tutoring were provided. This underscores the importance of teaching interventions accompanying destreaming initiatives, especially in mathematics.

Bavis reported on a similar destreaming initiative at Evanston Township High School in Illinois, USA.[117] A large and diverse public school, Evanston destreamed English, biology and history classes in 2010. Like the Rockville case, Evanston accompanied destreaming with unification of the curriculum. Whereas previously, four streams were each taught different curricula, following destreaming, all students were taught the curriculum previously reserved for the top (honours) stream. Importantly, efforts were also made to develop more effective pedagogy:

> Teams of English, history, and biology teachers developed the detracked courses. In each subject, teachers identified the skills students would need in junior and senior AP courses and how teachers might develop those skills in freshman year.
>
> — P. Bavis.[118]

A more systematic and structured approach to assessment, characterised by formative feedback, also accompanied the initiative.

After five years of destreaming, the school registered an American College Testing (ACT)[119] score of 23.9 – the highest in the school's history and nearly four points higher than the national average. Furthermore, more students attained college-ready scores in Advanced Placement (AP)[120] exams than had occurred previously. The gains were greatest for students of mixed-ethnicity (23 percentage points) and Asian students (19 percentage points). Gains for Black students (6 percentage points), Hispanic students (8 percentage points) and low-income students (7 percentage points) were more modest, and their overall attainment remained well below that of Asian and White students. Thus, while all student demographics improved their attainment under destreaming, educational inequality – the gaps between demographics with previously lower and higher attainment – actually got larger.

Like the Rockville initiative, destreaming at Evanston was accompanied by the provision of additional resources. The school provided study centres before, during, and after school on every weekday, so that students could consult teachers. A homework centre was established for four days per week and academic support was offered on some Saturdays. Furthermore, an academic intervention team was implemented, to identify and assist struggling students.

Neither of the Rockville or Evanston interventions can be taken as generalisable evidence in favour of destreaming. Uncontrolled variables may, partly or wholly, explain the improvements. Furthermore, there were confounds built into both. One would expect the provision of additional resources to improve attainment, irrespective of destreaming. The provision of a challenging curriculum to all students and, in the case of Evanston, the provision of frequent formative feedback, is also likely to have played a part. All these additional resources, and changes to curriculum and assessment, could have been enacted without destreaming.

The success of the Rockville and Evanston initiatives contrasts with another such intervention in the San Francisco Unified School District.[121] In the 2014–15 school year, the district abolished accelerated middle- and high-school mathematics classes and replaced them with unstreamed grouping for algebra in 9th grade (Year 10) and geometry in 10th grade (Year 11).

Loveless presented data from California's Smarter Balanced assessments that showed some improvement for ethnic demographics that had previously performed well – White and Asian students.[122] However, scores for Black students – who had the poorest scores before destreaming – were essentially unchanged, and scores for Hispanic students deteriorated. While the gap between White and Black students and between White and Hispanic students also worsened across California – including in districts that still stream – the gaps widened by

more than average in the Unified School District. Destreaming in the Unified School District did nothing to reduce educational inequality, in fact, it amplified it.

As Loveless noted, the caveats that apply to the Rockville and Evanston initiatives also apply to the Unified School District initiative; the findings are not generalisable and may be attributable to extraneous variables. Nonetheless, the different outcomes of the Rockville and Evanston initiatives on one hand, and the Unified School District initiative on the other, illustrate factors that are likely to be critical in the success of any destreaming initiative. Rockville and Evanston implemented the curriculum previously used for their top streams for all students and put in place supplementary classes and other resources to support lower-attainment students. The Unified School District, however, abolished the most advanced curriculum, and there is no indication that they offered any additional resources. This suggests that the support and resourcing that accompany destreaming initiatives are crucial to their efficacy.

According to Rubin, "The most powerful examples of detracking occur in schools where changes in instruction, institutional structures, and beliefs occur simultaneously to support the academic success of all students in newly challenging and stimulating settings".[123] The examples we have reviewed here support Rubin's claim. It remains an open question whether the improvements in attainment achieved in the Rockville and Evanston districts could have been achieved by putting the additional classes and resources in place, without abolishing streaming.

CHAPTER 4
Discussion and recommendations

The international evidence, especially evidence from meta-analyses, suggests that average attainment is usually similar in streamed and unstreamed environments. Nonetheless, streaming is often associated with greater than average disparities in learning, with greater attainment for higher-attainment streams cancelled by lesser attainment for lower-attainment streams. It is not clear, however, that streaming itself is the *cause* of that issue.

Gamoran noted that lower streams are often given less engaging curricula and pedagogy than higher streams.[124] When destreaming is not accompanied by redressing these issues, it is usually unsuccessful in raising attainment or reducing educational inequality, as illustrated by Loveless' Unified School District case study.[125] Many of the curriculum, pedagogical and assessment practices that typically accompany successful destreaming initiatives are educationally effective in their own right, as Hattie's meta-analysis showed.[126] Most could also be enacted in a streamed environment. So, again, the extent to which the apparent success of these destreaming initiatives is directly attributable to destreaming itself, is unclear.

Causal relationships can be reliably established only with experimental research designs. Slavin noted that experimental and correlational studies tend to differ in their findings, with the latter being more likely to associate increased learning disparities with streaming.[127] Slavin cited several experimental studies from the mid-twentieth century, in which students were randomly assigned to streamed or unstreamed environments. He noted that this research revealed no discernible impact on attainment gaps between students with lower and higher prior attainment.[128]

In experimental studies, grouping by prior attainment is set up as an experimental manipulation for the purpose of the research. Factors like curriculum, pedagogy and prior attainment are usually held constant across the streamed and unstreamed conditions. Correlational studies usually focus on already-existing streamed classroom environments, often with the kinds of associated curricular and pedagogical deficits for lower streams that Gamoran noted.[129]

The learning trajectories of students in different streams are likely to vary irrespective of streaming itself. Whether students are streamed or not, attainment gaps tend to increase over time according to the well-known "Matthew effect".[130] Therefore, if prior attainment is not accounted for, ongoing differences in rates of attainment cannot be taken as reliable evidence that streaming causes attainment gaps to increase.

One possible mediator of the effect of streaming on lower-attainment students is their separation from higher achieving students in their classrooms, who could otherwise support their learning.[131] This may work through direct peer support, differences in classroom culture, or effects on teacher expectations. In streamed schools, students in lower streams are deprived of interaction with higher-attaining peers. This is a potential mediator of the higher attainment gaps often observed in streamed environments.

Stereotype threat

If streaming directly impacts educational equality, it is likely to be at least partly due to a phenomenon known as *stereotype threat*.

This explanation would predict being placed in a low stream to be especially deleterious for students from demographic groups with lower-than-average educational attainment.

Stereotype threat describes an effect of negative stereotypes on members of stereotyped groups. When a member of a group is reminded of a stereotype attributed to that group, their behaviour or performance can come to conform with it. Stereotype threat was first observed by Steele and Aronson in relation to the academic performance of African American undergraduate students.[132] In this research, African American and White students were each divided into two groups, each of which was given a standardised test. One group was told that it was a genuine test of verbal ability. The other was told that the test was being carried out to understand psychological factors in solving verbal problems. African American students in the former group scored less well in the test than those in the latter. White students' performance was unaffected by the different information about the nature of the test. Steele and Aronson theorised that, when African American students were told that their verbal skills were being tested, the stereotype that African Americans have poor verbal skills was activated, causing them to lose confidence and motivation.

Hartley and Sutton showed a similar effect in relation to sex stereotypes.[133] They administered reading, writing and mathematics assessments to primary-aged boys and girls. One group were told that boys tend to perform less well at school than girls, while another group were told just to do as well as they could. Boys in the former group did less well than boys in the latter group, while girls' performance was the same in both conditions.

Stereotype threat is a risk of streaming. Any student placed in a lower stream, especially over successive years, may be vulnerable to poor learning efficacy. Stereotype threat exacerbates that vulnerability for students from demographics stereotyped as being less academically capable. In New Zealand that includes Māori and Pasifika and, in co-educational schools, boys. Bavis[134] noted that Claude Steele, one of the researchers who originally demonstrated stereotype threat has shown that "feedback grounded in high standards … reduces stereotype threat in students of color".[135] This suggests that the enriched curricula and attention given to formative assessment in many successful destreaming initiatives may be instrumental in the reduced educational disparities between ethnic groups often observed following these initiatives.

We noted the research of Card and Giuliano who showed that the establishment of a special class for gifted students was beneficial for them without causing any deterioration in the attainment of other students.[136] This finding lends some support to the hypothesis that the exacerbation of attainment gaps often associated with streaming, is attributable to stereotype threat. The placing of a few students – recognised as being unusually able – in a separate class, seems less likely to activate negative learning stereotypes than being placed in a class oneself that is, explicitly or implicitly, known to be for 'low achievers.'

Many students have higher attainment in some subjects than others. A system under which students can be placed in different streams for different subjects would help to ensure that they experience appropriate curriculum and pedagogy in all subjects. When students are grouped within, rather than between, classes, learning can be enhanced in mathematics[137] and early reading[138] without exacerbating educational inequality. In the second-order meta-analysis of Steenbergen-Hu et al. within-class grouping was positively associated with achievement for students at all levels of prior attainment.[139]

Separating students within classes, or for just some subjects, may make lower-attainment students feel less singled out than placing them in low-streamed classes, potentially reducing stereotype threat. And when grouping takes place within classes, students with lower attainment can still accrue benefit from the presence of higher attainment students in the same classes.

Cognitive load

The argument that grouping students by prior attainment enables teachers to better target their pedagogy is supported by cognitive load theory.[140] The cognitive load of a task is the extent to which the mental processes associated with performing that task occupy the resources of attention and working memory. Working memory is a short-term human memory system that stores information while we consciously use and manipulate it. It is heavily involved in learning new skills and knowledge.

Cognitive load must be managed by adapting the pace of teaching and learning. Because working memory has a very limited capacity and because its contents decay quickly unless actively maintained, tasks drawing on its resources are effortful and require concentration. When the demands of a learning task exceed the capacity of working memory, students experience cognitive overload. Cognitive overload causes feelings of confusion and, if it persists, frustration and demotivation.

With practice, tasks that are initially mediated by working memory are encoded in long-term memory. At that point, they become cognitively automatic, meaning that concentration is no longer required to perform them, and they no longer consume working memory resources. Likewise, with rehearsal, knowledge becomes encoded in long-term memory and available for automatic retrieval.

When progress in an area of learning depends on prior knowledge and skill, that prior learning must be stored in long-term memory before new learning that builds on it is approached. If it is not, cognitive overload is inevitable because both the prerequisite material and the new material must then be processed simultaneously in working memory. Teachers must pace learning carefully, to ensure that a student has encoded prerequisite knowledge in long-term memory before undertaking new learning that depends on it.

Not all learning places the same demands on working memory. Mathematics and early literacy are especially demanding because the knowledge and skills required to make progress are particularly hierarchical. In these learning areas, then, it is essential for teachers to ensure that their students automatise each step. That is easier to do if there is not too much variation in the curriculum stage students have reached within a class. If students are grouped according to prior learning, this variation is reduced. For that reason, the advantages of streaming in high cognitive load areas like early reading and mathematics may outweigh any deleterious effects of the practice, especially if steps are taken to mitigate those effects.

Summary

Stereotype threat and cognitive load theory provide differing accounts of the likely impact of streaming on educational achievement. Stereotype threat might account for findings that streaming impacts negatively on students placed in low streams, especially students in groups that are stereotyped as struggling at school. On the other hand, cognitive load theory provides a rationale for differentiating pedagogy for students at different levels of attainment. Both must be taken seriously, whether streaming is used or not.

In schools in which streaming is used, the risk of stereotype threat might be minimised by encouraging growth mindset, regular review of streaming placements, and rigorous use of formative feedback. In schools that do not stream, other ways must be found to minimise cognitive overload, especially in cognitively demanding curriculum areas like mathematics and early literacy. Flexible within-class grouping, and other ways of differentiating instruction, should be explored. The risk of cognitive overload might be especially great in schools that destream, given that many of their teachers will not be used to catering to students at very different stages of curriculum progress within the same classes.

Even if streaming is, on balance, harmful to educational equality, an outright national ban may be more harmful still. As comparing the case studies of Burris and Wellner,[141] Bavis[142] and Loveless[143] illustrate, destreaming is more likely to be successful when it is accompanied by modifications to curriculum, pedagogy and assessment. Successful destreaming initiatives have usually also involved supplementary classes or workshops for students previously in lower-attainment groups. Schools that undertake destreaming voluntarily can plan for and resource these elements. But if every school in New Zealand that uses streaming suddenly had to cease the practice, most would not be able to undertake the required steps to ensure that destreaming was successful. Furthermore, attempting to force destreaming on unwilling schools and communities would likely result in 'streaming by stealth' and result in poor implementation of destreaming.

A better approach would be to carry out a detailed, large scale, quantitative study on the effects of various streaming practices, publish the results, and provide resources to support schools to adapt their practices in light of that evidence.

Recommendations

Rather than seeking to ban streaming, the Ministry of Education should commission research on its prevalence, nature and educational effects in New Zealand. That research should then be published, and schools encouraged to adapt their practice according to the evidence it provides. That may entail cessation or modification of streaming practices.

From the point of view of investigating the causal effects of streaming on educational attainment, the ideal approach would be experimental, with students randomly allocated to streamed and unstreamed environments. However, this approach may be impractical in New Zealand's environment of self-governing schools. That leaves the statistical control of extraneous variables as the best available option. Fortunately, New Zealand's integrated database stores very detailed socioeconomic, educational and other data on every New Zealander. This resource would enable researchers to model the effects of streaming on education, accounting for a wide range of other variables.

The core of a study on streaming should be comparison of educational progress over a school year of students in different class arrangements. The starting point would be a Ministry survey of all New Zealand schools, to determine the type and extent of attainment grouping in each. This would include the school years during which students are streamed; which subjects are streamed; whether students are placed in the same stream for all subjects or different streams for different subjects; the numbers of separate streams; and the frequency with which students' placement in streams is reviewed. Within-class grouping practices should be probed by this survey. The survey should also gather information on the prevalence of formative feedback and other pedagogical practices that are known to be effective.

Following a national survey of streaming, samples of schools enacting various kinds of streaming, as well as ones that do not stream, should be selected. Students should be assessed at the beginning and end of a school year in key curriculum areas to measure progress. At primary level, this would include reading, writing and numeracy. At secondary level, it would include major subject areas. Students should also complete psychometric instruments to measure their learning efficacy, assimilation of negative stereotypes and locus of control in respect of their learning.

Analyses should condition progress on a range of socioeconomic data from the integrated database. The indicators used to calculate the equity index[144] for the purposes of school funding would suffice. Prior attainment, measured by the initial assessment, should also be a conditioning variable.

The differential impact of different streaming practices (including no streaming) on students at different levels of curriculum advancement could then be compared, accounting for these covariates. Impact on both learning progress and the psychometric constructs should be analysed.

The results of the study should be published in a form that is readily interpretable by schools and the public. If streaming is shown to be an ineffective, inequitable or harmful practice by such research, it is likely to provide a strong impetus for schools to discontinue it, especially if resources are provided to assist them adapt their practices based on the research. This is likely to be a more effective approach and get more schools on board with a need to change, than an outright ban.

Endnotes

1 Education Act 1877 (41 Victoriae 1877 No 21).

2 T Loveless, *How Well Are American Students Learning? With Sections on the Latest International Tests, Tracking and Ability Grouping, and Advanced Math in 8th Grade* (The Brown Center on Education Policy, 2013).

3 S. Hallam, J. Ireson and J. Davies. "Primary pupils' experiences of different types of grouping in school," *British Educational Research Journal* 30(4) (2004), 515–533; S. Parsons and S. Hallam. "The impact of streaming on attainment at age seven: evidence from the Millennium Cohort Study," *Oxford Review of Education*, 40(5) (2014), 567–589.

4 R.E. Slavin. "Ability grouping and student achievement in elementary schools: A best- evidence synthesis," *Review of Educational Research,* 57 (1987), 347–350; R.E. Slavin. "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis," *Review of Educational Research,* 60(3) (1990), 471–499.

5 R. Marks. *Ability Grouping in Primary Schools* (UK: Critical Publishing, 2016).

6 OECD. "Sorting and selecting students between and within schools" in *PISA 2018 Results (Volume V): Effective Policies, Successful Schools* (Paris: OECD Publishing, 2020).

7 S. Parsons, and S. Hallam. "The impact of streaming on attainment at age seven: evidence from the Millennium Cohort Study," op. cit.; S. Steenbergen-Hu, M. Makel, M. and P. Olszewski-Kubilius. "What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K-12 Students' Academic Achievement: Findings of Two Second-Order Meta-Analyses," *Review of Educational Research*, 86(4) (2016), 849-899.

8 Institute of Education Sciences. *Pre-COVID Ability Grouping in U.S. Public School Classrooms* (National Centre for Education Statistics at IES, 2021. Website.

9 See, for example, L. Lu. "The market phenomenon in Taiwanese junior high-level education: Behind the battles between streaming and mixed-ability grouping," *Comparative Education*, 48(3) (2012), 367-383.

10 Tokona Te Raki: Māori Futures Collective. *Ending Streaming in Aotearoa.* 2021. Website.; Tokona Te Raki: Māori Futures Collective. *Kōkirihia*. 2022. Website.

11 BERL et al. *He Awa Ara Rau: A Journey of Many Paths*. 2019. Website.; Tokona Te Raki. *Ending Streaming in Aotearoa*, op.cit.; Tokona Te Raki. *Kōkirihia*, op. cit.

12 L. Kenny. (2022, March 23). "Abolish school streaming to improve education for Māori students, says research". *Stuff*. Website.

13 D. Pomeroy, D., K.L. Jones, M. Azarmandi and S. Tolbert. "Could academic streaming in New Zealand schools be on the way out? The evidence suggests it should be" (2020, September 11). *The Conversation*. Website.

14 Ministry of Education. *Briefing Note: Update on work on streaming.* Version date 2021, May 12th, p.1. Website.

15 A. Chmielewski. "An International Comparison of Achievement Inequality in Within- and Between-School Tracking Systems," *American Journal of Education*, 120(3) (2014), 293-324; OECD. "Sorting and selecting students between and within schools" op. cit.

16 OECD. "Sorting and selecting students between and within schools", op. cit.

17 Study in Germany. *German Education System.* Website.

18 NZQA. *University Entrance* (2023, Oct 12). Website.

19 S. McNaughton, N. Rosedale, W. Fraser and J. Oldehaver. *Enhancement of Tertiary Education Strategy Priority Frameworks*. (The University of Auckland, 2014).

20 NZQA. *NCEA Pathways through Senior Secondary*. 2023. Website.

21 S. Jensen, I. Madjar and E. McKinley. "Students and NCEA course choices and allocations," *Set: Research Information for Teachers,* (2) (2010), 45–51. doi:10.18296/set.0439.

22 Ibid. 46.

23 OECD. "Selecting and Grouping Students," in *What Makes Schools Successful? Resources, Policies and Practices – Volume IV* (Paris: OECD Publishing, 2013); OECD. "Sorting and selecting students between and within schools", op. cit.

24 H. Noble, and J. Smith. "Issues of validity and reliability in qualitative research," *Evidence-Based Nursing*, 18 (2015), 34-35, 34.

25 Tokona Te Raki. *Ending Streaming in Aotearoa*, op. cit.

26 Tokona Te Raki. *Kōkirihia*, op. cit.

27 Tokona Te Raki: Mō Kā Uri A Muri Ake Nei. n.d. Website.

28 Tokona Te Raki: About Us. n.d. Website.

29 NZQA. *NCEA Pathways through Senior Secondary*, op. cit.; Tokona Te Raki. *Ending Streaming in Aotearoa*, op.cit.; Tokona Te Raki. *Ending Streaming in Aotearoa*, op. cit.

30 See, for example, R. Bishop, R., M. Berryman and J. Wearmouth. *Te Kotahitanga: Towards effective education reform for indigenous and other minoritised students* (NZCER Press, 2014).

31 See, for example, M. Alansari, M. Hunia and J. Eyre. *A rapid review of racism in schools: Working paper* (NZCER, 2020).

32 See, for example, K. Eriksson, J. Lindvall, O. Helenius and A. Ryve. "Socioeconomic Status as a Multidimensional Predictor of Student Achievement in 77 Societies," *Frontiers in Education,* 6 (2021). https://doi.org/10.3389/feduc.2021.731634

33 The Treasury. *Statistical Analysis of Ethnic Wage Gaps in New Zealand - Analytical Paper 18/03.* 2018. Website.

34 J. Haar. (2022, 11 August). Interview by Breakfast, *One News*. Website. https://www.1news.co.nz/2022/08/11/expert-fears-maori-unemployment-rate-dropping-for-perverse-reason/

35 Tokona Te Raki. *Kōkirihia*, op. cit.

36 Ibid. 18, italic added.

37 W. Elley. "To Stream or Not to Stream? Research on streaming: a review," *Set: Research Information for Teachers*, 2 (1976); A. Gamoran. *Tracking and Inequality: New Directions for Research and Practice. Working Paper.* (Wisconsin Center for Education Research, 2009); S. Glock, S. Krolak-Schwerdt and I. Pit-Ten Cate. "Are school placement recommendations accurate? The effect of students' ethnicity on teachers' judgments and recognition memory," *European Journal of Psychology of Education* 30(2) (2015), 169–188; J. Ireson, S Hallam and C Hurley. "What are the effects of ability grouping on GCSE attainment?" *British Educational Research Journal*, 31 (2005), 443–458; E.A. Hanushek and L. Woessmann. "Does Educational Tracking Affect Performance and Inequality? Differences in-Differences Evidence Across Countries," *The Economic Journal*, 116 (2006), 63-76; J. Hattie. *Visible learning: A synthesis of over 800 meta-analyses related to achievement*. (London: Routledge, 2009); C.C. Burris and D.T. Garrity. *Detracking for excellence and equity* (Association for Supervision and Curriculum Development (ASCD), 2008); A. Datnow and V. Park. (2018). *Professional Collaboration with Purpose Teacher Learning Towards Equitable and Excellent Schools.* (New York: Routledge, 2018).

38 W. Elley "To Stream or Not to Stream? Research on streaming: a review," op. cit.

39 A. Gamoran. *Tracking and Inequality: New Directions for Research and Practice. Working Paper,* op. cit.

40 J Hattie. *Visible learning: A synthesis of over 800 meta-analyses related to achievement,* op. cit.

41 E.A. Hanushek and L. Woessmann. "Does Educational Tracking Affect Performance and Inequality? Differences in-Differences Evidence Across Countries," op. cit.

42 J. Ireson et al. "What are the effects of ability grouping on GCSE attainment?" op. cit.

43 W. Elley "To Stream or Not to Stream? Research on streaming: a review," op. cit.

44 A. Gamoran. *Tracking and Inequality: New Directions for Research and Practice. Working Paper*, op. cit.

45 J Hattie. *Visible learning: A synthesis of over 800 meta-analyses related to achievement*, op. cit.

46 R.E. Slavin. "Ability grouping and student achievement in elementary schools: A best- evidence synthesis," op. cit.

47 C.M. Connor, F. J Morrison, B. J. Fishman, C. Schatschneider, and P. Underwood. "The early years: Algorithm-guided individualized reading instruction," *Science*, 315 (2007), 464–465; C.M. Connor, S. B. Piasta, B Fishman, S. Glasney, C. Schatschneider, E. Crowe, P. Underwood and F.J. Morrison. "Individualizing Student Instruction Precisely: Effects of Child by Instruction Interactions on First Graders' Literacy Development," *Child Development*, 80(1) (2009), 77-100.

48 W. Elley "To Stream or Not to Stream? Research on streaming: a review," op. cit. 8.

49 A. Gamoran. *Tracking and Inequality: New Directions for Research and Practice. Working Paper*, op. cit.

50 S. Glock et al. "Are school placement recommendations accurate? The effect of students' ethnicity on teachers' judgments and recognition memory," op. cit.

51 W. Elley "To Stream or Not to Stream? Research on streaming: a review," op. cit.; A. Gamoran. *Tracking and Inequality: New Directions for Research and Practice. Working Paper*, op. cit.; J Hattie. *Visible learning: A synthesis of over 800 meta-analyses related to achievement,* op. cit.

52 C.M. Broaded. "The limits and possibilities of tracking: Some evidence from Taiwan," *Sociology of Education*, 70 (1997) 36–53.

53 H. Ayalon and A. Gamoron. (2000). "Stratification in Academic Secondary Programs and Educational Inequality in Israel and the United States," *Comparative Education Review,* 44(1) (2000), 54-80.

54 A. Gamoran. "Curriculum standardization and equality of opportunity in Scottish secondary education, 1984–1990," *Sociology of Education*, 29 (1996), 1–21.

55 W. Elley "To Stream or Not to Stream? Research on streaming: a review," op. cit.

56 J Hattie. *Visible learning: A synthesis of over 800 meta-analyses related to achievement*, op. cit. 91.

57 A. Gamoran. "Curriculum standardization and equality of opportunity in Scottish secondary education, 1984–1990," op. cit.

58 See, M. Johnston and S. Martin. (2023). *Who Teaches the Teachers? Reforming Initial Teacher Education in New Zealand* (Wellington: New Zealand Initiative, 2023). Website.

59 W. Elley "To Stream or Not to Stream? Research on streaming: a review," op. cit.

60 R.E. Slavin. "Ability grouping and student achievement in elementary schools: A best- evidence synthesis," op. cit.

61 C.M. Connor et al. "The early years: Algorithm-guided individualized reading instruction," op. cit.; C.M. Connor et al "Individualizing Student Instruction Precisely: Effects of Child by Instruction Interactions on First Graders' Literacy Development," op. cit.

62 W. Elley "To Stream or Not to Stream? Research on streaming: a review," op. cit.; J Hattie. *Visible learning: A synthesis of over 800 meta-analyses related to achievement*, op. cit.

63 Ministry of Education. *He Whakaaro: Does Streaming Work?* 2021. Website.

64 G. Anthony. and R. Hunter, R. "Grouping Practices in New Zealand Mathematics Classrooms: Where Are We at and Where Should We Be?" *New Zealand Journal of Educational Studies* 52, (2017) 73-92.

65 H. Turner et al. "Teacher expectations, ethnicity and the achievement gap," op. cit.

66 S. Glock et al. "Are school placement recommendations accurate? The effect of students' ethnicity on teachers' judgments and recognition memory," op. cit.

67 G. Hornby and C. Witte. "Ability Grouping in New Zealand High Schools: Are Practices Evidence-Based?" *Preventing School Failure*, 58(2) (2014), 90-95.

68 G. Anthony and R. Hunter, R. "Grouping Practices in New Zealand Mathematics Classrooms: Where Are We at and Where Should We Be?" op. cit.

69 Ibid.

70 S. Glock et al. "Are school placement recommendations accurate? The effect of students' ethnicity on teachers' judgments and recognition memory," op. cit.

71 H. Turner et al. "Teacher expectations, ethnicity and the achievement gap," op. cit.

72 Tokona Te Raki. *Ending Streaming in Aotearoa*, op. cit.

73 OECD. "Selecting and Grouping Students," op. cit.

74 S. Martin. "Ability grouping for mathematics classes, 15-year olds', from PISA 2012 Results: What Makes Schools Successful?" in *Resources, Policies and Practices (Volume IV)* (Paris: OECD Publishing (2016).

75 OECD. "Sorting and selecting students between and within schools," op. cit.

76 V. Vandenberghe. "Evaluating the magnitude and the stakes of peer effect's analysing science and math achievement across OECD," *Applied Economics*, 34 (2002), 1283-1290; R. Benito, M. Alegre and I. Gonzalez-Balletbo. (2015). "School Segregation and Its Effects on Educational Equality and Efficiency in 16 OECD Comprehensive School Systems," *Comparative Education Review*, 58(1) (2015), 104-134.

77 V. Vandenberghe. "Evaluating the magnitude and the stakes of peer effect's analysing science and math achievement across OECD," op. cit.

78 R. Benito et al. (2015). "School Segregation and Its Effects on Educational Equality and Efficiency in 16 OECD Comprehensive School Systems," op. cit.

79 N. Mons. *Les Nouvelles Politiques Éducatives: La France Fait-elle Les Bons Choix?* (Paris: Presses Universitaires de France, 2007).

80 R. Benito et al. (2015). "School Segregation and Its Effects on Educational Equality and Efficiency in 16 OECD Comprehensive School Systems," op. cit.

81 T. Agasisti and S. Longobardi. "Educational institutions, resources, and students' resiliency: an empirical study about OECD countries, *Economics bulletin* 34(2) (2014), 1055-1067; T. Agasisti and P. Zoido. "Comparing the Efficiency of Schools Through International Benchmarking: Results from an Empirical Analysis of OECD PISA 2012 Data". *Educational Researcher* 47(6) (2018), 352–362.

82 J Hattie. *Visible learning: A synthesis of over 800 meta-analyses related to achievement*, op. cit. 91.

83 R. E. Slavin. "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis," op. cit.

84 S. Steenbergen-Hu et al. "What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K-12 Students' Academic Achievement: Findings of Two Second-Order Meta-Analyses," op. cit.

85 Rui, N. "Four decades of research on the effects of detracking reform: Where do we stand? -Asystematic review of the evidence," *Journal of Evidence-Based Medicine,* 2(3) (2009), 164-183.

86 J Hattie. *Visible learning: A synthesis of over 800 meta-analyses related to achievement,* op. cit. 91.

87 R. E. Slavin. "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis," op. cit.

88 Rui, N. "Four decades of research on the effects of detracking reform: Where do we stand? -Asystematic review of the evidence," op. cit.

89 Ibid.

90 S. Steenbergen-Hu et al. "What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K-12 Students' Academic Achievement: Findings of Two Second-Order Meta-Analyses," op. cit.

91 J. Ireson, J. and S Hallam. "Raising Standards: Is Ability Grouping the Answer?" *Oxford Review of Education*, 25(3) (1999), 343–358, 347. https://www.jstor.org/stable/1050925

92 S. Parsons and S. Hallam. "The impact of streaming on attainment at age seven: evidence from the Millennium Cohort Study," op. cit.

93 J. Trinidad and R. King. (2022). "Ability grouping predicts inequality, not achievement gains in Philippine schools: findings from PISA 2018," *Educational Research for Policy and Practice*. 21(2) (2022), 305-322.

94 Ibid.

95 Hong, C. Corter, Y. Hong and J. Pelletier. "Differential Effects of Literacy Instruction Time and Homogeneous Ability Grouping in Kindergarten Classrooms: Who Will Benefit? Who Will Suffer?" *Educational Evaluation and Policy Analysis*, 34 (1) (2012), 69-88.

96 L. Linchevski and B Kutscher. "Tell Me with Whom You're Learning, and I'll Tell You How Much You've Learned: Mixed Ability versus Same-Ability Grouping in Mathematics," *Journal for Research in Mathematics Education*, 29(5) (1998), 533-554.

97 A. Kerckhoff. "Effects of Ability Grouping in British Secondary Schools," *American Sociological Review*, 51(6) (1986), 842-858.

98 C. Cheung and E. Rudowicz. "Academic Outcomes of Ability Grouping among Junior High School Students in Hong Kong," *The Journal of Educational Research,* 96(4) (2003), 241-254.

99 R.E. Slavin. "Ability grouping and student achievement in elementary schools: A best- evidence synthesis," op. cit.

100 Borg, W. "Ability grouping in the public schools," *The Journal of Experimental Education*, 34 (1965),1-97.

101 R.E. Slavin. "Ability grouping and student achievement in elementary schools: A best- evidence synthesis," op. cit.

102 R.E. Slavin. "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis," op. cit.

103 H. Forgasz. "Streaming for Mathematics in Years 7-10 in Victoria: An Issue of Equity?" *Mathematics Education Research Journal*, 22(1) (2010), 57-90.

104 A. Gamoran. "The stratification of high school learning opportunities," *Sociology of Education*, 60 (1987), 135-155.

105 C. M. Connor et al. "The early years: Algorithm-guided individualized reading instruction," op. cit.

106 See, for example, G. Hong et al. "Differential Effects of Literacy Instruction Time and Homogeneous Ability Grouping in Kindergarten Classrooms: Who Will Benefit? Who Will Suffer?" op. cit.

107 G.W. Thompson. "The Effects of Ability Grouping upon Achievement in Eleventh Grade American History", *The Journal of Experimental Education*, 42:4 (1974), 76-79.

108 Card, D. and L. Giuliano. "Can Tracking Raise the Test Scores of High-Ability Minority Students?" *American Economic Review*, 106(10) (2016), 2783–2816.

109 B. Ballis and Heath, K. (2021). "The Long-Run Impacts of Special Education," *American Economic Journal: Economic Policy*, 13(4) (2012), 72–111.

110 Ibid.

111 Card, D. and L. Giuliano. "Can Tracking Raise the Test Scores of High-Ability Minority Students?" op. cit.

112 B. Ballis and Heath, K. (2021). "The Long-Run Impacts of Special Education," op. cit.

113 C. Burris and K. Welner. "Closing The Achievement Gap by Detracking," *Phi Delta Kappan*, 86(8) (2005), 562-640; P. Bavis. "Detracked – And going strong," *Phi Delta Kappa International*, 98(4) (2016), 37-42.

114 T. Loveless. (2022, March 13). "San Francisco's Detracking Experiment". Tom Loveless Blog. Website. https://tomloveless.com/posts/san-franciscos-detracking-experiment/

115 C. Burris and K. Welner. "Closing The Achievement Gap by Detracking," op. cit.

116 Ibid. 596.

117 P. Bavis "Detracked – And going strong," op. cit.

118 Ibid. 38.

119 The ACT is an entrance exam for college admissions in the US.

120 AP refers to college-level courses offered to high school students in the US.

121 T. Loveless. (2022, March 13). "San Francisco's Detracking Experiment," op. cit.

122 Ibid.

123 Rubin, B. "Tracking and Detracking: Debates, Evidence, and Best Practices for a Heterogeneous World," *Theory into Practice*, 45(1) (2006), 4-14, 5.

124 A. Gamoran. *Tracking and Inequality: New Directions for Research and Practice. Working Paper*, op. cit.

125 T. Loveless. (2022, March 13). "San Francisco's Detracking Experiment," op. cit.

126 J Hattie. *Visible learning: A synthesis of over 800 meta-analyses related to achievement*, op. cit. 91.

127 R.E. Slavin. "Ability grouping and student achievement in elementary schools: A best- evidence synthesis," op. cit.

128 Ibid.

129 A. Gamoran. *Tracking and Inequality: New Directions for Research and Practice. Working Paper*, op. cit.

130 Merton, R. K. "The Matthew Effect in Science," *Science*, 159(3810) (1968), 56–63.

131 V. Vandenberghe. "Evaluating the magnitude and the stakes of peer effect's analysing science and math achievement across OECD," op. cit.

132 C.M. Steele and J. Aronson. "Stereotype threat and the intellectual test performance of African Americans," *Journal of Personality and Social Psychology*. 69(5) 1995), 797–811. DOI:10.1037/0022-3514.69.5.797

133 B. L. Hartley and R.M. Sutton. (2013). "A stereotype threat account of boys' academic underachievement," *Child Development*, 84(5) (2013), 1716–1733. https://doi.org/10.1111/cdev.12079

134 P. Bavis "Detracked – And going strong," op. cit. 38.

135 C.M. Steele and J. Aronson. "Stereotype threat and the intellectual test performance of African Americans," op. cit. 38.

136 Card, D. and L. Giuliano. "Can Tracking Raise the Test Scores of High-Ability Minority Students?" op. cit.

137 Borg, W. "Ability grouping in the public schools," op. cit.

138 R.E. Slavin. "Ability grouping and student achievement in elementary schools: A best- evidence synthesis," op. cit.

139 S. Steenbergen-Hu et al. "What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K-12 Students' Academic Achievement: Findings of Two Second-Order Meta-Analyses," op. cit.

140 For a detailed discussion, see: J. Sweller, P. Ayers, and S. Kalyuga. *Cognitive Load Theory* (Springer 2011). DOI 10.1007/978-1-4419-8126-4

141 C. Burris and K. Welner. "Closing The Achievement Gap by Detracking," op. cit.

142 P. Bavis "Detracked – And going strong," op. cit.

143 T. Loveless. (2022, March 13). "San Francisco's Detracking Experiment," op. cit.

144 Ministry of Education. *The equity Index.* 2023. Website.

References

Agasisti, T. and S. Longobardi. "Educational institutions, resources, and students' resiliency: an empirical study about OECD countries, *Economics bulletin* 34(2) (2014), 1055–1067.

Agasisti, T. and P. Zoido. "Comparing the Efficiency of Schools Through International Benchmarking: Results from an Empirical Analysis of OECD PISA 2012 Data". *Educational Researcher* 47(6) (2018), 352–362.

Alansari, M., M. Hunia and J. Eyre. *A rapid review of racism in schools: Working paper* (NZCER, 2020).

Anthony, G. and R. Hunter, R. "Grouping Practices in New Zealand Mathematics Classrooms: Where Are We at and Where Should We Be?" *New Zealand Journal of Educational Studies* 52, (2017) 73–92.

Ayalon, H. and A. Gamoron. (2000). "Stratification in Academic Secondary Programs and Educational Inequality in Israel and the United States," *Comparative Education Review,* 44(1) (2000), 54–80.

Ballis, B. and Heath, K. (2021). "The Long-Run Impacts of Special Education," *American Economic Journal: Economic Policy,* 13(4) (2012), 72–111.

Bavis, P. "Detracked – And going strong," *Phi Delta Kappa International*, 98(4) (2016), 37–42.

Benito, R., M. Alegre and I. Gonzalez-Balletbo. (2015). "School Segregation and Its Effects on Educational Equality and Efficiency in 16 OECD Comprehensive School Systems," *Comparative Education Review*, 58(1) (2015), 104–134.

Bishop, R., M. Berryman and J. Wearmouth. *Te Kotahitanga: Towards effective education reform for indigenous and other minoritised students* (NZCER Press, 2014).

Borg, W. "Ability grouping in the public schools," *The Journal of Experimental Education*, 34 (1965), 1–97.

Broaded, C. M. "The limits and possibilities of tracking: Some evidence from Taiwan," *Sociology of Education*, 70 (1997) 36–53.

Burris, C.C. and D.T. Garrity. *Detracking for excellence and equity* (Association for Supervision and Curriculum Development (ASCD), 2008).

Burris, C. and K. Welner. "Closing The Achievement Gap by Detracking," *Phi Delta Kappan*, 86(8) (2005), 562–640.

Business and Economic Research Ltd (BERL), Waikato-Tainui, Southern Initiative, Te Rūnanga o Ngāi Tahu. *He Awa Ara Rau: A Journey of Many Paths*. 2019. Website. https://knowledgeauckland.org.nz/media/1902/he-awa-ara-rau-a-journey-of-many-paths-berl-tsi-et-al-nov-2019.pdf

Card, D. and L. Giuliano. "Can Tracking Raise the Test Scores of High-Ability Minority Students?" *American Economic Review*, 106(10) (2016), 2783–2816.

Cheung, C. and E. Rudowicz. "Academic Outcomes of Ability Grouping among Junior High School Students in Hong Kong," *The Journal of Educational Research*, 96(4) (2003), 241–254.

Chmielewski, A. "An International Comparison of Achievement Inequality in Within- and Between-School Tracking Systems," *American Journal of Education*, 120(3) (2014), 293–324.

Connor, C. M., F. J Morrison, B. J. Fishman, C. Schatschneider, and P. Underwood. "The early years: Algorithm-guided individualized reading instruction," *Science*, 315 (2007), 464–465.

Connor, C.M., S. B. Piasta, B Fishman, S. Glasney, C. Schatschneider, E. Crowe, P. Underwood and F.J. Morrison. "Individualizing Student Instruction Precisely: Effects of Child by Instruction Interactions on First Graders' Literacy Development," *Child Development*, 80(1) (2009), 77–100.

Datnow, A. and V. Park. (2018). *Professional Collaboration with Purpose Teacher Learning Towards Equitable and Excellent Schools*. (New York: Routledge, 2018).

Education Act 1877 (41 Victoriae 1877 No 21). New Zealand Legal Information Institute (NZLII). Website. http://www.nzlii.org/nz/legis/hist_act/ea187741v1877n21224/

Elley, W. "To Stream or Not To Stream? Research on streaming: a review," *Set: Research Information for Teachers,* 2 (1976).

Eriksson, K., J. Lindvall, O. Helenius and A. Ryve. "Socioeconomic Status as a Multidimensional Predictor of Student Achievement in 77 Societies," *Frontiers in Education*, 6 (2021). https://doi.org/10.3389/feduc.2021.731634

Forgasz, H. "Streaming for Mathematics in Years 7-10 in Victoria: An Issue of Equity?" *Mathematics Education Research Journal*, 22(1) (2010), 57–90.

Gamoran, A. "The stratification of high school learning opportunities," *Sociology of Education*, 60 (1987), 135–155.

——— "Curriculum standardization and equality of opportunity in Scottish secondary education, 1984–1990," *Sociology of Education*, 29 (1996), 1–21.

——— *Tracking and Inequality: New Directions for Research and Practice. Working Paper*. (Wisconsin Center for Education Research, 2009).

Glock, S., S. Krolak-Schwerdt and I. Pit-Ten Cate. "Are school placement recommendations accurate? The effect of students' ethnicity on teachers' judgments and recognition memory," *European Journal of Psychology of Education* 30(2) (2015), 169–188.

Haar, J. (2022, 11 August). Interview by Breakfast, *One News*. Website. https://www.1news.co.nz/2022/08/11/expert-fears-maori-unemployment-rate-dropping-for-perverse-reason/

Hallam, S., J. Ireson and J. Davies. "Primary pupils' experiences of different types of grouping in school," *British Educational Research Journal* 30(4) (2004), 515–533.

Hartley, B. L. and R.M. Sutton. (2013). "A stereotype threat account of boys' academic underachievement," *Child Development*, 84(5) (2013), 1716–1733. https://doi.org/10.1111/cdev.12079

Hanushek, E.A. and L. Woessmann. "Does Educational Tracking Affect Performance and Inequality? Differences in-Differences Evidence Across Countries," The *Economic Journal*, 116 (2006), 63-76.

Hattie, J. *Visible learning: A synthesis of over 800 meta-analyses related to achievement*. (London: Routledge, 2009).

Hong, G., C. Corter, Y. Hong and J. Pelletier. "Differential Effects of Literacy Instruction Time and Homogeneous Ability Grouping in Kindergarten Classrooms: Who Will Benefit? Who Will Suffer?" *Educational Evaluation and Policy Analysis*, 34 (1) (2012), 69–88.

Hornby, G., and C. Witte. "Ability Grouping in New Zealand High Schools: Are Practices Evidence-Based?" *Preventing School Failure,* 58(2) (2014), 90–95.

Institute of Education Sciences. *Pre-COVID Ability Grouping in U.S. Public School Classrooms* (National Centre for Education Statistics at IES, 2021. Website. https://nces.ed.gov/pubs2021/2021139.pdf

Ireson, J. and S Hallam. "Raising Standards: Is Ability Grouping the Answer?" *Oxford Review of Education*, 25(3) (1999), 343–358. http://www.jstor.org/stable/1050925

Ireson, J., S Hallam and C Hurley. "What are the effects of ability grouping on GCSE attainment?" *British Educational Research Journal*, 31 (2005), 443–458.

Jensen, S., E. McKinley, E. and I. Madjar. "NCEA subject choices in mid—low-decile schools: What schools and parents need to know about the university pathway," *Set: Research Information for Teachers*, 2, (2010), 37–44.

Johnston, M. and S. Martin. (2023). *Who Teaches the Teachers? Reforming Initial Teacher Education in New Zealand* (Wellington: New Zealand Initiative, 2023). Website. https://www.nzinitiative.org.nz/reports-and-media/reports/who-teaches-the-teachers/

Kenny, L. (2022, March 23). "Abolish school streaming to improve education for Māori students, says research". *Stuff*. Website. https://www.stuff.co.nz/national/education/128116534/abolish-school-streaming-to-improve-education-for-mori-students-says-research

Kerckhoff, A. "Effects of Ability Grouping in British Secondary Schools," *American Sociological Review*, 51(6) (1986), 842–858.

Linchevski, L. and B Kutscher. "Tell Me with Whom You're Learning, and I'll Tell You How Much You've Learned: Mixed Ability versus Same-Ability Grouping in Mathematics," *Journal for Research in Mathematics Education,* 29(5) (1998), 533–554.

Loveless T. (2022, March 13). "San Francisco's Detracking Experiment". Tom Loveless Blog. Website. https://tomloveless.com/posts/san-franciscos-detracking-experiment/

——— *How Well Are American Students Learning? With Sections on the Latest International Tests, Tracking and Ability Grouping, and Advanced Math in 8th Grade* (The Brown Center on Education Policy, 2013).

Lu, L. "The market phenomenon in Taiwanese junior high-level education: Behind the battles between streaming and mixed-ability grouping," *Comparative Education*, 48(3) (2012), 367–383.

Marks, R. *Ability Grouping in Primary Schools* (UK: Critical Publishing, 2016).

Martin, S. "Ability grouping for mathematics classes, 15-year olds', from PISA 2012 Results: What Makes Schools Successful?" in *Resources, Policies and Practices (Volume IV)* (Paris: OECD Publishing (2016).

McNaughton, S., N. Rosedale, W. Fraser and J. Oldehaver. *Enhancement of Tertiary Education Strategy Priority Frameworks*. (The University of Auckland, 2014).

Merton, R. K. "The Matthew Effect in Science," *Science*, 159(3810) (1968), 56–63.

Ministry of Education. *Briefing Note: Update on work on streaming*. Version date 2021, May 12th. Website. https://www.education.govt.nz/assets/Documents/our-work/information-releases/Advice-Seen-by-our-Ministers/April-2021/R-3-1255301-Update-on-Work-on-Streaming_Redacted.pdf

———. *He Whakaaro: Does Streaming Work A review of the evidence?* Insights Education, October 2021. Website. https://www.educationcounts.govt.nz/__data/assets/pdf_file/0012/210027/He-Whakaaro-Does-streaming-work-.pdf

———. *The Equity Index*. Website. https://www.education.govt.nz/our-work/changes-in-education/equity-index/

Mons, N. *Les Nouvelles Politiques Éducatives: La France Fait-elle Les Bons Choix?* (Paris: Presses Universitaires de France, 2007).

Noble, H. and J. Smith. "Issues of validity and reliability in qualitative research," *Evidence-Based Nursing*, 18 (2015), 34–35.

NZQA. *NCEA Pathways through Senior Secondary.* 2023. Website. https://www2.nzqa.govt.nz/about-us/publications/insights-papers/ncea-pathways-through-senior-secondary/#:~:text=M%C4%81ori%20students%20are%20twice%20as,school%20as%20NZ%20European%20students.

———. *University Entrance* (2023, Oct 12). Website. https://www.nzqa.govt.nz/qualifications-standards/awards/university-entrance/

OECD. "Selecting and Grouping Students," in *What Makes Schools Successful? Resources, Policies and Practices – Volume IV* (Paris: OECD Publishing, 2013).

———. "Sorting and selecting students between and within schools" in *PISA 2018 Results (Volume V): Effective Policies, Successful Schools* (Paris: OECD Publishing, 2020).

Parsons, S. and S. Hallam. "The impact of streaming on attainment at age seven: evidence from the Millennium Cohort Study," *Oxford Review of Education*, 40(5) (2014), 567–589.

Pomeroy, D., K.L. Jones, M. Azarmandi and S Tolbert. "Could academic streaming in New Zealand schools be on the way out? The evidence suggests it should be" (2020, September 11). *The Conversation*. Website. https://theconversation.com/could-academic-streaming-in-new-zealand-schools-be-on-the-way-out-the-evidence-suggests-it-should-be-145617

Rubin, B. "Tracking and Detracking: Debates, Evidence, and Best Practices for a Heterogeneous World," *Theory into Practice,* 45(1) (2006), 4–14.

Rui, N. "Four decades of research on the effects of detracking reform: Where do we stand? -Asystematic review of the evidence," *Journal of Evidence-Based Medicine*, 2(3) (2009), 164–183.

Slavin, R. E. "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis," *Review of Educational Research*, 60(3) (1990), 471–499.

——— "Ability grouping and student achievement in elementary schools: A best- evidence synthesis," *Review of Educational Research*, 57 (1987), 347–350.

Steele, C.M. and J. Aronson. "Stereotype threat and the intellectual test performance of African Americans," *Journal of Personality and Social Psychology.* 69(5) 1995), 797–811. DOI:10.1037/0022-3514.69.5.797.

Steenbergen-Hu, S., M. Makel, M. and P. Olszewski-Kubilius. "What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K-12 Students' Academic Achievement: Findings of Two Second-Order Meta-Analyses," *Review of Educational Research*, 86(4) (2016), 849–899.

Study in Germany. *German Education System.* Website. https://www.studying-in-germany.org/german-education-system/

Sweller, J., P. Ayers, and S. Kalyuga. *Cognitive Load Theory* (Springer 2011). DOI 10.1007/978-1-4419-8126-4

The Treasury. *Statistical Analysis of Ethnic Wage Gaps in New Zealand - Analytical Paper 18/03.* 2018. Website: https://www.treasury.govt.nz/sites/default/files/2018-08/ap18-03.pdf

Thompson, G.W. "The Effects of Ability Grouping upon Achievement in Eleventh Grade American History", *The Journal of Experimental Education*, 42:4 (1974), 76–79.

Tokona Te Raki: Māori Futures Collective. *Ending Streaming in Aotearoa*. 2021. Website. https://www.growwaitaha.co.nz/media/2173/ending-streaming-in-aotearoa-final.pdf

———. *Kōkirihia*. 2022. Website. https://www.maorifutures.co.nz/research/research-item-one/

———. Mō Kā Uri A Muri Ake Nei. Website. n.d. https://www.maorifutures.co.nz/

——— About Us. Website. n.d. https://www.maorifutures.co.nz/about-us/

Trinidad, J. and R. King. (2022). "Ability grouping predicts inequality, not achievement gains in Philippine schools: findings from PISA 2018," *Educational Research for Policy and Practice.* 21(2) (2022), 305–322.

Turner, H., C. Rubie-Davies and M. Webber. "Teacher expectations, ethnicity and the achievement gap," *New Zealand Journal of Education Studies,* 50 (2015), 55–69.

Vandenberghe, V. "Evaluating the magnitude and the stakes of peer effect's analysing science and math achievement across OECD," *Applied Economics*, 34 (2002), 1283–1290.

Streaming – the practice of separating students into different classes based on their prior attainment – is widely used in New Zealand secondary schools. Some commentators have recently called for a ban, arguing that streaming harms the educational achievement of students placed in lower streams, especially Māori.

In *Class Divides* we review the evidence on the educational effects of streaming. We discuss risks associated with the practice, as well as challenges associated with discontinuing it.

To date, there has been no large-scale quantitative study investigating the effects of streaming on student achievement in the New Zealand context. We call for such a study to be commissioned by the Ministry of Education, to inform national policy and school practice.

$25.00

ISBN
978-1-7386009-8-4 (print)
978-1-7386009-9-1 (online)

THE NEW ZEALAND INITIATIVE

www.nzinitiative.org.nz

The New Zealand Initiative
PO Box 10147
Wellington 6143