Bangor University

DOCTOR OF PHILOSOPHY

'More bang for your buck': Exploring the use of health economics in an educational context.

Tiesteel, Emma

Award date:
2024

Awarding institution:
Bangor University

Link to publication

Prifysgol Bangor

Bangor University

'More bang for your buck': Exploring the use of health economics in an educational context.

Emma Tiesteel

Thesis submitted to the School of Medical and Health Sciences, Bangor University, in partial fulfilment for the degree of Doctor of Philosophy

May 2023

Declaration

Yr wyf drwy hyn yn datgan mai canlyniad fy ymchwil fy hun yw'r thesis hwn, ac eithrio lle nodir yn wahanol. Caiff ffynonellau eraill eu cydnabod gan droednodiadau yn rhoi cyfeiriadau eglur. Nid yw sylwedd y gwaith hwn wedi cael ei dderbyn o'r blaen ar gyfer unrhyw radd, ac nid yw'n cael ei gyflwyno ar yr un pryd mewn ymgeisiaeth am unrhyw radd oni bai ei fod, fel y cytunwyd gan y Brifysgol, am gymwysterau deuol cymeradwy.

I hereby declare that this thesis is the results of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.

Yellow Slip

Contact the Dyslexia Team
for further details if necessary
or go to
https://www.bangor.ac.uk/studentservices/disability/documents/sep-20-marking-guidelines-eng.pdf
for marking guidelines

Acknowledgements

I would like to offer sincere gratitude to my supervisors, Professor Carl Hughes, Dr Richard Watkins and Dr Carys Jones. Although the journey has not been a linear one, they have been able to guide me through the difficult landscape of being a researcher, analysist, and a writer. While no one could predict the unprecedented effect of the pandemic on the education system, I am grateful to my supervisors giving me the time, space, and support to come out the other end.

I would sincerely like to thank the participants who took part in the research. For the teachers who took time out of their busy days to conduct interviews. And to the pupils, who were brilliantly honest, and candid with their opinions and experiences.

I offer the biggest high five to Bethany and Kaydee, you both have been a source of support throughout (the good, the bad and the ugly). While you both had your own mountains to climb, you both were able to down tools when I needed support.

To Stacy, I thank you for your unfiltered advice, support and understanding. For always believing in me, through the thick and thin. You know what makes me tick and without you I would not be the woman I am today. I give thanks to you Kat, for always being there when I need to rant, cry or laugh. For convincing me to try new things and to push at my boundaries. And Winners, from humble undergrads to almost academics, thank you for always being there, even though you are a brilliantly formidable woman, your kindness and gentleness with me will never be forgotten.

I offer the greatest thanks to my parents Peter and Sharon, while I was not the model child in school, this PhD would not be possible without your love, support, and babysitting. To my brothers and sister, thank you for always understanding that I'm busy, never putting any pressure on me. I adore you all, in different ways.

To my children, gosh, it's all for you. Tyler, Lil Ems, Neo, Nate and Lou Lou, you won't know this, but you keep me going. I am in awe of every single one of you, how you face your own challenges pushes me to get through my own. I want to thank you all for understanding and getting me through the tough times by making me laugh, debate and giving me a few more grey hairs. Keep going kids, the world is your oyster.

And finally, to the longest suffering husband there ever was. Shaun, no words can even express my love and gratitude, for everything you do for me and this family. Your selflessness knows no bounds, and because of this I have been able to aim for the sky. I dedicate this thesis to you, because without you it would not be possible. I love you more, the end. I win.

Contents Page

List of tables

List of figures

Abstract

Background:

There is a changing paradigm in education towards the creation of a more evidence informed system to improve outcomes for learners (Gorard et al., 2020). There is also a drive to consider cost effective policy making, or activities in education that offer value for money in terms of public spending (Levin, 2001; OECD, 2022; Owen et al., 2022), but considerations are needed, the drive for efficiency can lead to narrow views of what is effective and risks reducing education to a small set of variables (Levin, 2012). With spending on education in the UK totalling £116 billion (Drayton et al., 2022), there is a growing need for countries to demonstrate efficiency and equality within the education provision they provide. However, there is still a need to gather evidence of what tools/activities in education are effective in allowing us to evaluate: (a) whether provision is value for money, and (b) how to allocate provision equitably to support learner's outcomes that is in addition to having good quality teachers.

Health economics is an established evaluation framework that facilitates decision makers particularly in the UK by using economic evaluations (Buxton, 2006; Corbacho & Pinto- Parades, 2012). This thesis utilises an economic evaluation framework traditionally used in health economics and evaluates if the framework can be used in an education setting. The purpose of this thesis is to explore the feasibility of using health economics in education to improve the quality of evaluation of a large-scale regional school improvement programme. This thesis also aims to explore what are the barriers and/or facilitators to the use of health economic approaches in education. With the aim to open a more constructive discussion about what activities are effective, and the consideration around measuring outcomes.

Structure

Chapter 1 outlines relevant literature in this field alongside the rationale for this thesis, whilst Chapter 2 reviews the associated research design and methods used throughout the empirical chapters.

The focus of the systematic review (Chapter 3) is on the existing literature published on the use of economic evaluations in school settings. By homing the search in on five evaluation methods typically used in health economic, we captured literature with clear cost and outcome data. A narrative synthesis was conducted on the included studies, as well as a

critical reflection on the methods used. This allowed a data driven approach to consider outcome measures and importantly the correct cost framework that set the groundwork for the following chapters.

Chapter 4 presents two further empirical studies; reporting the qualitative data that we collected as part of the evaluation framework for a regional formative assessment intervention project (FAIP). Chapter 5 (study 4) presents the full economic evaluation of the FAIP, including an evaluation of survey data and a full economic cost, including sensitivity analyses, and a summary of the qualitative data.

Chapter 6 explains the methodological framework required when using cost evaluations in education, including difficulties migrating a framework from health research to education and the difficulties that researchers may face when using cost analysis in education. This chapter also discusses the theoretical underpinnings used in health economics and builds on the work that has already been done by recommending further research in this field.

Chapter 7 draws together the main findings from each of the studies and discusses the utility of using health economics methods in education, including some recommendations for future directions in both policy and research.

Results

The systematic review (Chapter 3) identified 12 published economic evaluations conducted within educational settings. This led to recommendations related to collecting cost information at an early stage, using comparable outcome indicators, and establishing the comparability of populations. Furthermore, cost-effectiveness ratios should be presented (where feasible), and there is a need for a critical appraisal tool for education settings to be developed. The cost per pupil for FAIP, was £67.55 (inflated to 2022-23 prices) and no statistically significant differences were found on all quantitative indicators. Thematic analysis of teacher and pupil interviews identified some important, positive features of FAIP. Teachers identified socially inclusive classroom, reduced workloads, improved standards of work and being able to adjust instruction based on the pupils needs. The pupils articulated their understanding of the strategies, and the outcomes.

Conclusion

This thesis demonstrates that economic evaluation methods can be applied to an education setting, but some considerations are needed around how costs are collected and calculated. This thesis identified how economic evaluations are used in education, and critically evaluated how they differed from health economic frameworks. Further research is needed around the identification of centralised costs; the provision for training for education

researchers on economic methods; and the integration of economic evaluations into the journey for evidence. The thesis concludes with a theoretical consideration of how to measure educational value to develop the theory of economics in an educational context.

COVID-19 Statement

The first national lockdown coincided with the beginning of my third year. All data for Chapters 3, 4, and 5 was collected prior to the first lockdown period. One study had to be omitted from the thesis due to difficulty collecting follow up data (discussed in Chapter 7). During lockdown periods and subsequent homeworking restriction, supervisory support was delivered through online Microsoft Teams. The sustained impact to the education system, meant supporting my children to continue their education, impacting the progress of the thesis. The continued impact on the thesis resulted in the funded period ending and having to take up employment. The supervisory team have supported the write up and analysis throughout.

Glossary

Chillies: Indicator used to differentiate classroom tasks.

CHU-9D: A health-related quality of life questionnaire.

Cost- Benefit Analysis: Calculating all the benefits and cost of an intervention and converting into monitory values.

Cost Consequence Analysis: Calculating the incremental costs of an intervention and identifying benefits and presenting the results without aggravating the costs and benefits.

Cost Minimization Analysis: Where the outcomes are expected to be the same, the analysis centres on the least costly inputs to achieve the same outcomes.

Cost Utility analysis: The costs are measured in monitory values and the outcomes in utility gained.

Cost-effectiveness Analysis: Costs and outcomes of one intervention are compared to alternating interventions.

Cost-effectiveness plane: A visual representation of the intervention compared to alternatives in cost-effectiveness analysis.

EEF: Education Endowment Foundation

e-FSM: Eligible for Free School Meals.

Extra Welfarism: Focuses on the measurement of health rather that utility, using methods to elicit preferences of health states.

Incremental cost: The cost of an additional unit of output.

Incremental Cost-effectiveness ratio (ICER) The cost per additional unit of output compared to the alternative.

Learning zone diagram: A visual aid for pupils to identify different zones of learning (from the centre: comfort zone, learning zone and panic zone).

Marginal Costs: The cost of making or producing one additional unit.

OECD: Organisation for Economic Co-operation and Development.

Opportunity cost: The costs of the forgone benefits by choosing an alternative.

PISA: Programme for International Student Assessment.

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

PROSPERO: International prospective register of systematic reviews.

PSSRU: Personal Social Services Research Unit.

QALY: Quality Adjusted Life Years.

Sensitivity analysis: Varying the key parameters of the intervention, to test assumptions, this can enhance the robustness of the findings.

Utility: Within health economics this refers to the preference of health state given by an individual or society.

Welfare economics: The focus is around how allocation of resources impact social welfare.

Chapter 1 Background and context to the thesis.

"Getting value for money is going to be critical to schools' ability to secure improvements in pupil outcomes." (EEF, 2018, p10)

1.1 Statement of the problem

There is a changing paradigm in education towards the creation of a more evidence informed system to improve outcomes for learners (Gorard et al., 2020). This has been exemplified over recent years at school- and policy-making levels as evidenced by Welsh Government's National Strategy for Educational Research and Enquiry (NSERE) (Welsh Government, 2021) and an increase interest in adopting evidence-based provision in schools (Pegram et al., 2022; Gorard et al., 2020). These initiatives aim to inform policy and practice, and, in doing so, support the education professionals to utilise the best available evidence to improve learner outcomes. Being more evidence informed means understanding how we can best support learners to achieve the best outcome from their education (i.e., understanding 'what works' [Nelson & Campbell, 2017]).

The UK-based 'Education Endowment Foundation' (EEF) (Gov.UK, 2023) further support the evidence informed agenda in schools. The EEF support schools to access available evidence via jargon-free summaries of available evidence. There is also a drive to consider cost effective policy making, or activities in education that offer value for money in terms of public spending (Levin, 2001; OECD, 2022; Owen et al., 2022). Whilst policymakers and schools can make use of available summaries from organisations such as the EEF, there is little guidance on how much different courses of action cost in education. Some commentators in education suggest failing to provide information on cost limits the scope for decision makers (Hummel-Rossi & Ashdown, 2002; Kraft, 2020; Levin, 2001; Levin & Benfield, 2015).

Many areas of public spending and education face a difficult financial landscape. In real terms, Sibieta (2020) highlighted that spending per pupil has decreased by 5% in Wales and by 8% in England and Northern Ireland between 2011-2019. Moreover, schools were increasingly providing services outside of the academic focus (Morris & Dobson, 2019). With spending on education in the UK totalling £116 billion (Drayton et al., 2022), there is a growing need for countries to demonstrate efficiency and equality within the education

provision they provide. Understanding how to get value for money in education is seen as the way to improve outcomes for pupils and education systems around the world (Agasisti, 2014; OECD, 2022). However, there is still a need to gather evidence of what tools/activities in education are effective in allowing us to evaluate (a) whether provision is value for money and (b) how to allocate provision equitably to support learner's outcomes.

Looking to other fields that focus on evaluating public spending, there is a more developed framework used in health services to assess the costs and outcomes of different causes of action. Health economics is an established evaluation framework that facilitates decision makers particularly in the UK by using economic evaluations (Buxton, 2006; Corbacho & Pinto- Parades, 2012). While there are no formal frameworks in education that assess the value for money in the UK, the use of robust and tested economic evaluation methods could support a move to a more joined up approach to spending and outcomes in education. This approach could, in turn, support decision makers to delegate funding to effective courses of action within the budgetary frameworks they have. This thesis utilises an economic evaluation framework traditionally used in health economics and evaluates if the framework can be used in an education setting.

1.1.2 Background and rationale of this thesis

Previous research has identified a need within the education system to make more evidence informed decisions and improve the use of more promising approaches in schools. However, to better inform decision makers we need data around the cost and effectiveness of the different strategies/programmes used in schools. This can support evaluation within the context of value for money and improve the quality of decision making by policymakers. As such, public money will be spent more effectively, in line with a value for money approach (Hummel-Rossi & Ashdown, 2002; OECD, 2022).

Cost analysis has been discussed as an important next step to a more evidence-informed education system (Detrich, 2020; Owen et al., 2022). Despite an increased focus on evidence informed practise within the literature, there is little work around the cost implication that can limit the scalability of effective interventions. Moreover, very little research has reported the associated cost analysis for interventions with a strong evidence base (Scammacca et al., 2020). Although many systems provide comprehensive state funded education to all school age learners, there remain embedded educational inequalities that are a

cause for concern for education ministers and policymakers around the world (Banerjee, 2016).

1.1.3 The attainment gap

Perhaps the most pertinent issue in education is the attainment gap. This is the qualification difference between pupils from disadvantaged backgrounds and their peers from non-disadvantaged backgrounds. Robertson (2021) explained that this is also known as the poverty related attainment gap. Deprivation—particularly focusing on the relationship between socio-economic status and attainment—has a long and difficult history. Before the introduction of the 1944 Education Act, as few as 1% of working-class pupils went to grammar schools (Jackson & Marsden, 2012). The 1944 Education Act aimed to support working-class children to gain qualifications via investment in human capital. Contextually, at this time Britain needed more economic outputs as the government was trying to tackle the five giants of post war poverty (Want, Disease, Ignorance, Squalor, and Idleness; or in a modern-day terms, poverty, health, education, housing, and employment opportunities). However, this Act did little to close the attainment gap. Even the marketization of the education systems in the 1980s left large numbers of learners leaving the education system with little more than basic mathematics and English (Ball, 2016). Additionally, the selective pathways between grammar and comprehensive school after the high stakes testing of the 11+ exams did little to combat social mobility and the diversity of students in Tertiary education (Paterson & Iannelli, 2007).

Even in the modern context, the attainment gap has remained largely unchanged Analysis of GSCE data collected before the COVID-19 pandemic highlights that the attainment gap has changed little over the last ten years despite several further policy initiatives (2011- 2019). In Wales, pupils from disadvantaged backgrounds are 22-23 months behind their non-disadvantaged peers; this has only reduced from 24 months in 2011. The picture is similar in England with pupils from disadvantaged background being 18 months behind their non -disadvantaged pupils; down from 20 months in 2011 (Cardim-Dias & Sibieta, 2021). There has been some indication that pupils who are persistently disadvantaged (i.e., those who maintain e-FSM status throughout school) are less segregated into clusters of schools showing some improvements for this sub-set of disadvantaged pupils (Gorard et al., 2022).

There are persistent and intergenerational issues related to poor attainment which can create a long shadow of poverty through the lifespan. The intergenerational impact of poor attainment can have a cumulative impact and there is a strong correlation with poor housing, poor health, unstable employment, crime, and health harming behaviours (Andrews et al., 2017; Cardim-Dias & Sibieta, 2021; Hunt et al., 2021; Zajacova & Lawrence, 2018). Education is a right for all learners and can also be a preventative intervention; when done well, it can reduce health inequalities and decrease mortality rate (Solé-Auró & Alcañiz, 2016).

Education systems in the modern context generally follow meritocracy principles. That is, systems should give every learner an equal chance of gaining qualifications, foster talent and achievements, and support their career potential regardless of the socio-economic status of their families (Hing et al., 2011). All learners entering the education system should receive an equitable experience to develop their potential. However, this is not always the case, and this has led to focused educational policies and funding to support disadvantaged learners (Gorard, 2022).

The most relevant policy in Wales is the Pupil Development Grant (PDG). This was designed to support disadvantaged pupils and introduced by the devolved Welsh education administration in 2012 (Welsh Government, 2018). Other home nations have similar policies for equity funding, for example, in England the Pupil Premium (Gov. UK, 2021), in Scotland the Pupil Equity Fund (PEF; Education Scotland, 2022), and in Northern Ireland, the Pupil Premium Targeting Social Need (TSN) within the Common Funding Formula (CFF) (NIAO, 2021). These funds are awarded directly to schools to fund additional support and provision to improve outcomes for disadvantaged learners. In most cases these funds, fund additional provision on top of core provision. One of the many ways this is done is by purchasing additional interventions, which are delivered by teaching assistants to support the needs of disadvantaged learners. The current proxy measure of socio-economic disadvantage in the UK eligibility for free school meals (e-FSM). Free school meals are awarded to the children of families in receipt of welfare benefits.

In Wales, schools currently receive £1,150 PDG for each e-FSM pupil and those in care of the state (i.e., looked after children [LAC]; Welsh Government, 2018). In addition, schools receive funding for deprivation through local authority funding formulae. There is evidence that in some contexts targeting funding in this way fails to achieve the government policy outcome targets (Gorard et al., 2022; Sibieta, 2020).

Schools have relative autonomy on how they use their equity funding (PDG, PP, and PP+). However, there is some consensus that this money is often not being spent in effective and efficient ways (i.e., cost-effective). As such, the funding often does not reach those most in need of support (Gorard 2022; Read et al., 2020). For example, LAC and learners in key stage 4 (aged 14 – 16 years), constantly have disproportionately disappointing outcomes. The equity funding did show some small gains when first introduced over a decade ago, but analysis suggests a widening attainment gap after initial positive progress (Mannay & Lyttleton-Smith, 2019; Read et al., 2020). Estyn (2022) provide thematic reports exemplifying case studies of schools using the PDG effectively.

Whilst UK governments are pushing for equity within the education systems and providing additional funding for disadvantaged learners, there is no quantifiable data to demonstrate impact on attainment. In the UK, there is little evidence that targeted funding has yielded stable results for the attainment gap (Pye et al., 2015; Sibieta, 2020).

1.1.4 Has spending in education made a difference

On the global stage, an analysis of the educational spending in line with the Program International Student Assessment (PISA) from different countries, suggests that if the main outcome indicator is the PISA ranking then spending per head does not support the 'does money matter' argument. Figures suggest that Finland who spend on average $10,025 US dollars per head and rank 8th on the PISA assessments, were as the US spend on average $12,424 rank 24th (OECD, 2020). However, this is not a fair comparison due to some of the complex issues involved and the fact that education systems operate in very different contexts.

There is a wide range of research reporting how an increase in spending influences pupil outcomes. Jackson's (2020) literature review concluded that, in some instances, increases in spending has exogenous impacts. Looking to other nations, a recent meta-analysis of increased spending in the USA by Jackson and Mackevicious (2021) suggested that increasing spending by $1000 per pupil positively affects test scores by 0.03 standard deviations. In the UK, Gibbons et al. (2018) explored spending patterns within local authorities, modelling for disadvantage. They found that increasing spending at £1000 per pupil within a local authority level increased test scores by 0.30 standard deviations for pupils by the end of primary school.

To explore the issue of spending, educational researchers have used return on investment (ROI) evaluations to evaluate the benefits of early education. This has helped inform a policy move towards investing in early years education (Levin et al., 2017). Research on the High Scope Perry Preschool Program suggest the rate of return in 3:1 (Heckman et al., 2010) and programs like Head Start pass the cost-benefit test (Ludwig & Phillips, 2007). Reynolds and Temple (2008) published an in-depth review of cost-effective programs for use in early childhood. While early years education retains an evidence based economic argument, governments, policymakers, local authorities, and school leaders are yet to fully utilise the potential of such evaluations to provide informed answers to 'where money matters'.

There is research that supports the consensus that extra spending does have positive impacts (Gibbons et al, 2018; Jackson, 2020). However, little is known about what is effective outside of this research. Gorard (2022) suggested that there is still some work to identify effective and cost-effective allocations of funds to support learner outcomes. While the evidence suggests that more fiscal investment in education does yield positive outcomes, researchers are faced with exploring what works (educational research) and the financial impact (economic evaluations). The growing discourse within these communities is that economic evaluations are rare (Belfield & Levin, 2015; Levin, 2001). Without such research decision makers are not being fully informed (Hummel-Rossi & Ashdown, 2002; Machin et al., 2013).

1.1.5 Teacher quality/ agency and fidelity in the evidence-informed journey

While this thesis wants to explore 'what works' and 'how much' it cannot ignore the need for good quality teachers in the classroom. The importance of quality teaching is a matter of importance to government and is evident in policymaking. A review of initial teacher training in Wales highlighted the importance of having high-quality teachers and the need to be embedded in the initial teacher training (Welsh Government, 2015; Furlong et al. 2021). Research suggests that poorer quality teachers do not command the same outcomes for their pupils and poorer quality teachers are more prevalent in areas of disadvantage (Schleicher, 2018). To improve pupil outcomes, particularly for pupils from disadvantaged backgrounds, governments need to attract the right people, who are trained to a high standard and can be the best resource a school will have to support pupils (Rice, 2010; Fauth, et al. 2019). Good quality teachers engage in evidence-informed and evidence-based strategies,

these teachers are also able to implement strategies with fidelity, (Hill & Ericson, 2019) a good quality teacher can account for up to 30% of outcome variance for pupils (Hattie, 2003). As such good quality teachers need to be considered before other activities can be introduced in the classroom either through school level interventions or policy changes.

Teacher autonomy and school autonomy play a part in the impacts of the pedagogy and how schools and professionals utilize evidence from both within and outside of the settings. Schools that have a higher degree of autonomy can have a stronger impact on pupil outcomes. Building the capacity for autonomy at all levels of the school infrastructure supports schools to self-improve, this can be done by having quality teachers and leaders who can evaluate progress and make changes to practice to support learning. This autonomy is opposed to centralized top-down ways of working, and greater reliance on surveillance of schools and teaching teachers who have high demands on teacher from other activities not directly related to the classroom instruction limit the capacity for teachers to have the autonomy to develop evidence-based strategies within their practice (Larson, et al. 2018), allowing for autonomy for teacher improves there retention and satisfaction in the workplace (Worth & Van den Brande, 2020).

Attempts have been made to measure the quality of teaching particularly in America, where policymakers used Value Added Models (VAM) to assess teacher quality, this policy resulted in teachers losing their jobs if they did not produce standardized outcomes for their cohort. This high-stakes policy has seen many teachers leave the profession and legal challenges to the policy (Yeh, 2013). This surveillance and less autonomy on the teaching profession has a negative impact on the teaching profession and ultimately the pupils in the education system. In Japan, teachers are hired at a local authority level, and higher quality teachers are deployed to schools in disadvantaged areas (Schleicher, 2018). While this could be linked to pupils in Japan having higher international outcomes, it is outside the scope of this thesis. Good quality teaching should be the cornerstone of any education system, and the other policies, strategies, and activities should not only complement the work of teachers, but these activities must be evidence-informed and value for money.   Discussion within this thesis is a side from teachers' quality, it is concerned about the inputs either through policy (i.e. Pupil development grant) or individual interventions at the schools or county level (i.e. the region-wide formative assessment evaluated here).

1.2 Current changes in Welsh education

Wales has had a major reform journey since devolution in 2009, during which the Welsh Assembly were able to design education policies (Power, 2016). Evans (2021) detailed three distinct phases that the Welsh education system have been through since devolution:

During **the first phase** of reform following devolution Wales was able to set their own agenda for the educational policies. The changes saw a diversion from the systems and policies in England, focusing on the whole child rather than just academic capabilities. This was evidenced by the establishment of the 'Foundation Phase', whereby pupils were encouraged to learn through play and follows a more Scandinavian model of education. There was also a strong focus on the Welsh language and understanding of the sense of belonging to the county and the local community. Within this first phase power was decentralised and the rolling back of accountability measures, league tables were discontinued, and power was given to the local authorities who became responsible for raising standards (James & Colebourne, 2004). Wales wanted to move away from blanket testing and in 2008 teacher assessment were used to assess the pupils progress (OECD, 2014). Wales decided not to follow England's trend in encouraging the marketisation of education, for example outsourcing services, league tables, parental choice, and academisation. Wales also gradually phased out standardised testing for 11- and 14-year-olds.

Evans (2021) termed t**he second phase** as the accountability phase. This was enacted in response to the PISA rankings with growing concerns over Welsh pupils falling behind the rest of the UK (Rees & Taylor, 2015). The radical change in policy highlighted the need for government to drive up standards with a sharper focus on literacy and numeracy. A school categorisation system was developed, alongside national Reading and Numeracy tests for pupils in key stage 2 and 3 (7- to 14-year-olds). This policy shift aligned closely with the systems in the other home nation's (most notably England), whereby neo-liberal policy making was evident (Ball, 2016).

The third phase was prompted by the OECD (2014) *Improving schools in Wales* report. This led the way for a review of the education system in Wales by Professor Donaldson, who had previously supported the development of the Scottish curriculum. The report detailed 68 recommendations (Donaldson, 2015) and set in motion the development and roll out of the *Curriculum for Wales.* This curriculum is centred on four guiding principles:

- Ambitious, capable learners, ready to learn throughout their lives.

- Enterprising, creative contributors, ready to play a full part in life and work.

- Ethical, informed citizens of Wales and the world.

- Healthy, confident individuals, ready to lead fulfilling lives as valued members of society.

The other notable change was around six new areas of learning and experience (AoLEs) and literacy, numeracy and digital competency being cross-curricula responsibility. This development was seen as the corner stone of Welsh education policy (Evans, 2021), taming of the neoliberal policies. Perhaps most importantly, this led to the roll-out of a curriculum that is developed by schools and harnesses creative freedom as opposed to a dictated curriculum within a top-down model. A strength of this phase was that practitioners in the education system use the 'Pioneer' model to develop and disseminate the new curriculum. Pioneer teachers and schools are the close-to-practice knowledge brokers to the top of the education system (Kneen et al., 2021).

1.2.1 School Improvement service

While Evans (2021) provided a comprehensive analysis of the main points and changes in the education system since the devolution, one area that is not in the analysis was the development of the reginal consortia as part of the school improvement services. The school improvement journey was initiated as part of the Welsh Government's structural system changes to support and accelerate the school improvement model and the commitment to regional working in Wales (Estyn, 2022). The development of the regional consortia aimed to support the local authorities (LAs) in their statutory responsibilities to improve education for all learners. Their role was (and continues to be) to drive the reform journey and support the Welsh education system. Thus, the reginal consortia work with the LAs and schools to:

- improve learner outcomes for all young people,

- ensure the delivery of high-quality teaching and learning,

- support and empower school leaders to better lead their schools (Welsh Government, 2015, p5)

The regional consortia cover the different regions of Wales and represent groups of local authorities working together to deliver school improvement services. North Wales is covered by the North Wales regional school improvement service (GwE); Southeast Wales is

covered by Educational Achievement Service (EAS); Central and West Wales is covered by Mid Wales Education Partners; South West Wales is served by Partneriaeth; and, Central South Wales is covered by Central South Consortium (CSC). In March 2020, Neath Port Talbot withdrew from the consortia framework. and i not currently working with a regional consortium or a partnership.

The reginal consortia support education by providing support and challenge to schools. They also provide training for school leaders to ensure effective delivery of education across the system. By working in partnership with the LA and schools, they provide a lens of objectivity within an education system with the aim of fostering a self-improving system (Reynolds, 2015).

1.2.2 Why this is important

In Wales, the National Strategy for Educational Research and Enquiry (NSERE) (Welsh Government, 2021) is designed to support the creation of a more evidence-informed policy making framework alongside more evidence-informed practice at all levels of the education system. The Welsh Government are currently building a national infrastructure in Wales and working with key stakeholders in the UK and internationally to realise this aim (Welsh Government, 2021). To this end, the regional consortia are Welsh Government's key partners in the drive towards this goal.

The regional consortia are charged with "providing access to evidence of 'what works' in terms of closing gaps in attainment and support schools to implement and assess the impact of targeted intervention strategies". (Reynolds, 2015, p. 14). In Wales, the schools' inspectorate, Estyn, evaluates whether schools are using resources efficiently, and questions school leaders on the cost-effectiveness of the programmes and activities within the school (Estyn, 2020). While Estyn do provide thematic reports on effective practice, there is little evidence of quantifiable impact in these reports and limited conclusions around guidance for effective spending.

Bowden et al. (2020) argued that without any financial evaluation of spending in the broader terms in education decision makers are not fully informed. As such, they cannot make evidence informed decisions. School leaders need to be competent in their financial literacy but there lacks the professional development in this area. Financial stress is one of the biggest concerns that face school leaders (Scott, et al., 2021; Woods et al., 2022).

1.3 Exploring value for money

There is a need for schools to demonstrate effective use of their funding. Estyn (2022) have now built this into the inspectorate's evaluation frameworks. Additionally, there has also been a call for schools to have a better understanding of the areas in which increased spending is effective, that is addition to having quality teachers. Collectively, this information can support schools and policy makers in gathering the information needed to make informed decisions. Within education there are few system-wide frameworks that allow this type of analysis. Health research particularly the work of health economists is used to support the decision-making (Hoffmann et al., 2002; Frew & Breheny, 2020), so the focus of this thesis is to evaluate whether health economics techniques and frameworks can be used in an educational setting.

Spending on education comes second to spending on health. Whilst attempts have been made in education to investigate spending in the form of annual reports, they do not use economic evaluations that are more routinely used in healthcare (Farquharson et al., 2021). The focus of this thesis is to assess the potential to use these economic evaluation techniques and methodologies in education.

1.3.1 How will this be demonstrated

In 2017 GwE commenced the implementation of a regional Formative Assessment Implementation Project (FAIP) designed to improve teachers' skills. To test the feasibility of using health economics to evaluate a regional school improvement project, a research project was designed around an economic evaluation of FAIP (Chapter 2 explains the project in detail).

1.3.2 Policy context around Formative Assessment

The seminal work from Black and Wiliam (1998) placed formative assessment at the forefront of the educational agenda. Their influential publication *Inside the Black Box* (Black & Wiliam, 1998) highlighted the potential of formative assessment as an important school improvement strategy for policy makers and teachers. It has continued to form a central principle of education policy and teachers' continuing professional development (Wiliam, 2020). In 2008, the Labour Government's education policy emphasised the importance of

formative assessment in initial teacher education provision. Black and Wiliam (1998) argued that policies that promoted formative assessment (or Assessment for Learning [AfL] as it became known) were often under resourced with 'empty commitments' from senior administrators. A further impetus to change assessment policy using evidence from research came from the Assessment Reform Group's (ARG) publication, *'Inside the Black Box and Assessment for Learning: beyond the black box'* (Broardfoot et al., 1999). However, the context surrounding AfL policy, did not translate well within schools due to the backdrop of high stakes testing and league tables championed by the coalition government (James, 2011; 2017).

After a review of the Welsh Education system by Donaldson (2015), formative assessment was detailed as an assessment strategy and "as the bridge between teaching and learning" (Donaldson, 2015, p.76). Although not explicitly labelled as formative assessment, the new Curriculum for Wales guidance concludes:

"*This should be achieved by embedding assessment into day-to-day practice in a way that engages the learner and makes it indistinguishable from learning. This allows the practitioner to respond to the individual needs of the full range of learners within their classroom on an ongoing basis.*" (Welsh Government, 2019, p. 224)

Most of the research on large-scale formative assessment projects to improve pedagogy focus on teacher learning communities. For example, the King's-Medway-Oxfordshire Formative Assessment Project (KMOFAP) Black and Wiliam (2005) looked at how 36 teachers could develop formative assessment principles to improve outcomes for pupils. Teachers from different schools were given 25 different practical techniques that they could choose to implement in their classroom. Teachers attended full training days every five weeks and researchers were on-site to assess and support. Positive effects were identified using standardised test outcomes. However, Wiliam (2020) suggested that despite the positive outcomes for the pupils, this approach is not viable at scale or attractive to policymakers due to the time commitment and hands-on approach. To maximise formative assessment in this way, the researchers need to refine the design so that schools can implement it relatively independently with lower costs.

In another large-scale randomized control trial evaluated by the Education Endowment Fund (EEF), secondary school teachers received training and monthly workshops to trial and develop the use of formative assessment within the schools. This

revealed a positive result on the Attainment 8 GCSE scores, translating to two months progress (Speckesser et al, 2018). However, the effectiveness research for formative assessment has come under scrutiny (see Bennett, 2011; Dunn & Mulvenon, 2009). Effect sizes have been questioned and a meta-analysis of 42 independent effects sizes of formative assessment research suggests that the effect size is 0.20 rather than the 0.40 to 0.70 that was originally proposed (Kingston & Nash, 2011).

The need for further research has been identified, to improve the design and impact of development programs aimed at improving teachers' understanding of formative assessment principles (Kanjee & Mthembu, 2015). Little research has explored formative assessment from the perspective of pupils. Gaining an improved understanding from pupils on how formative assessment works in practice could support the design and implementation in the classroom (Cowie et al., 2018).

1.4 Economic frameworks

Economic evaluation, in its simplest form, is a balance sheet of cost and benefits (effects) of an intervention (Morris et al., 2012). There are 5 main types of economic evaluations that are common in health care analysis: cost-effectiveness analysis; cost-utility analysis; cost-benefit analysis; cost-minimisation; and cost-consequence analysis. These methods have a shared characteristic of the measurement and valuation of costs. However, they differ in their approach to the measurement and valuation of benefits.

1.4.1 Cost-effectiveness analysis

Cost-effectiveness analysis (CEA) is an economic framework that measures the outcome in natural units (e.g., life years gained, pain or symptom free days). CEA always uses a comparator even if this is usual care. The use of natural units allows disease specific outcomes to be compared against each other to identify whether an intervention is more effective in terms of costs and effects than a comparator. Incremental cost-effectiveness ratios (ICERs) can be used, informing decision makers what incremental effects and costs interventions have over each other. This is calculated by comparing two treatments in terms of costs and effects (i.e., by dividing the difference in costs and the effects; Morris et al., 2012).

1.4.2 Cost-utility analysis

Cost-utility analysis (CUA) is a type of economic evaluation is a form of cost-effectiveness analysis where the unit of effect is typically the Quality Adjusted Life Year (QALY), a utility-based measure which reports the effect of an intervention in terms of length and quality of life. It is used in evaluations where (1) improving health-related quality of life is a target of the intervention; and (2) the intervention may have a wide range of outcomes and the evaluators want to convert these into a single, generic, unit of outcome. QALYs are the most common type of utility measure used as they are internationally valued and accepted (Morris et al., 2012). QALYs are calculated using generic questionnaires (e.g. EuroQol [EQ]-5D or Health Utilities Index [HUI]) or condition specific measures (e.g., Asthma Quality of Life Questionnaire) condition specific questionnaires are mapped onto generic questionnaires (e.g., EQ-5D) to generate the QALY (Whitehead & Ali, 2010). The National Institute of Health and Care Excellence (NICE) uses a threshold of £20,000-£30,000 per QALY gained in health technology assessments to decide whether an intervention is cost-effective or not; except for end-of-life treatments, where a higher threshold is allowed (Collins & Latimer, 2013). Treatments or interventions that fall under the thresholds can be accepted into the health service and considered value for money (Towse, 2009).

1.4.3 Cost-benefit analysis and benefit-cost analysis

Cost-benefit analysis (CBA) and benefit-cost analysis (BCA)[1] value the outcomes and costs in monetary terms. By quantifying more than one benefit of an intervention, both these measures can give a better idea of the overall benefits of an intervention (rather than one domain such as cost-effectiveness analysis). Although this method has the benefit of being able to measure the tangible benefits of a program or intervention, it is not always possible to measure the intangible benefits (e.g., happiness and freedom from pain; Morris et al., 2012).

---

[1] Benefit-Cost Analysis (BCA) is the term more commonly used in education research, with the alternative term Cost-Benefit Analysis (CBA) used in health research. European researchers commonly adopt the term CBA, whereas in the United States the term BCA is used. However, they both refer to the same method and the term is used interchangeably by Andersson (2018). Other researchers, however, argue that CBA and BCA represent different moral perspectives and justification of decision (Zerbe, 2019). This review will take Andersson's (2018) view that they are the same method, and the term is used interchangeably.

Cost-benefit analysis produces a benefit-cost ratio (BCR). For example, a ratio of 3:1 would mean that for every £1 invested £3 is returned.

1.4.4 Cost-minimisation analysis

Economists use cost-minimisation analysis to identify the lowest cost option when an intervention and the alternative under consideration are expected to have the same or very similar outcomes (e.g., choosing between a branded drug verses a generic drug). Cost-minimisation analysis is not appropriate when the effectiveness of competing alternatives is not equivalent (Charles & Edwards, 2016).

1.4.5 Cost-consequence analysis

Cost-consequence analysis (CCA) lists costs and outcomes in a disaggregated way so that decision makers can choose the outcome that is the most important to them. CCA allows the decision maker to form their own opinion and prioritise which costs, and outcomes are important to their context (Charles & Edwards, 2016).

1.5 Purpose of the thesis

The purpose of this thesis is to explore the feasibility of using health economics in education to improve the quality of evaluation of a large-scale regional school improvement programme. This thesis also aims to explore what are the barriers and/or facilitators to the use of health economic approaches in education. To answer the feasibility question, we identified a large-scale school improvement programme focused on the delivery of formative assessment training for teachers based on a train the trainer model. A logic model was produced with stakeholders to assess all the components the research could cover and the expected outcomes. Data was collected using mixed methods and was quasi- experimental in design (discussed in greater detail in Chapters 3,4 and 5).

1.5.1 Thesis novel contribution to knowledge

The use of economic evaluations frameworks is underdeveloped. This thesis aims to improve our understanding of:

- The feasibility of using methods from health economics to build an economic evaluation framework for education.
- Evidence-based decision making in education by providing information on economic considerations.
- The potential barriers and facilitators to borrowing methods from one discipline to another.

1.5.2 Research Questions

The following questions are addressed by this thesis:

RQ1: What is the evidence for the effective use of health economics approaches to evaluate education programmes/provision?

RQ2: What is the impact of the Formative Assessment Implementation Project (FAIP), and does this represent value for money?

RQ3: What is the feasibility of using health economics approaches to evaluate a large-scale education programme in schools?

1.5.3 Wider generalisability

This thesis' primary aim is to explore using methods of health economics to evaluate activities in education, but the outcomes may generalise much wider. Firstly, this work will build on the limited economic insight to activities in education. In doing so, the growing literature can support educationalists at all levels of the education system to think systematically about spending patterns particularly in relation to outcomes. Decision makers can therefore access comprehensive information about models of evaluation which they can use to make informed decision around the strategies implemented within the education system. This is particularly important as the education system recovers from the global pandemic where there is a stronger focus on interventions that are effective and value for money.

1.5.4 Structure of the thesis

Chapter 1: Background and context of the thesis. This introduces the main concepts and the background to some of the issues faced in the education system. The chapter

discusses the main economic methods used in health along with literature on additional spending and the impact on learner outcomes. The research questions and the novel contribution are presented in this chapter. Figure 1.1 shows the structure of the thesis.

Chapter 2: Research design and methods. This chapter explains the research design and methods used to evaluate the regional formative assessment intervention project.

Chapter 3: A systematic review of economic evaluations in a school setting (Study 1). This focuses on the five economic methods used in health economics, with clear cost and outcome data. The time scale was not restricted to a particular period and no limitations were set for the type of interventions. For a study to be included in the review, they needed to have used an educational intervention with learners under 18-years-old, in a school setting and detail a cost per pupil. Once the screening had taken place, the following data was extracted in line with the systematic review protocol: Country, type of intervention, aims, population under study, type of economic evaluation, outcome measure, cost per pupil, inflated price to present day, perspective, main findings, and conclusion. The Drummond checklist was used to assess the quality of the included studies (Drummond et al., 2015). A narrative synthesis was conducted on the included studies, as well as a critical reflection on the methods used. This allowed a data driven approach to consider outcome measures and importantly the correct cost framework that set the groundwork for the following chapters.

Chapter 4: Interviews and focus groups (Study 2 and 3). This chapter details the qualitative data, that was collected as part of the evaluation framework for the regional formative assessment intervention project. Interviews were conducted with class teachers and focus groups with pupils in the intervention classes.

Chapter 5: Cost Consequence Analysis (Study 4). This chapter presents the full economic evaluation of the FAIP. This chapter includes the survey data (CHU-9D, SDQ, and QoLS) alongside national test data. We compared the data with that of a matched control group. This chapter details the full economic cost, sensitivity analysis, and a summary of the qualitative data.

Chapter 6: Methodological considerations: This chapter is an explanation of the methodological framework that is needed when using cost evaluations in education, including difficulties migrating a framework from health research to education and the difficulties that researchers may face when using cost analysis in education. This chapter discusses the theoretical underpinnings used in health economics and builds on the work that has already been done by recommending further research in this field.

Chapter 7: Discussion. This chapter draws together the main findings from each of the studies and discusses the utility of using health economics methods in education, including some recommendations for future directions in both policy and research.

***Figure 1.1Structure* of the thesis**

Chapter 1: Introduction

Research Question 1: What is the evidence for the effective use of health economics approaches to evaluate education programmes/provision?

Research Question 2: What is the effectiveness of a regional formative assessment programme, and does this represent value for money?

Research Question: What is the feasibility of using health economics approaches to evaluate a large-scale education programme?

Previous research in education.

Specifically looking at health economic methods in education.

Main outcome measures.

-S&D questionnaire
-CHU-9D Questionnaire
-School Satisfaction Questionnaire
-Focus groups
-Teacher interviews
-Attainment data

Chapter 2
Details the approach taken and methods used. The rational behind the data collection .

Chapter 3
A systematic review of evidence of health economics within the educational setting.

Chapter 4. 1 Focus group data

Chapter 4.2 Interview data

Chapter 5
CCA: Given the limited effectiveness data from quantitative methods. Cost consequence analysis will be performed.

Chapter 6
Methods paper: Discussion on how and what is needs to be consider to develop CEA in education.

Chapter 7: Discussion

Chapter 2 Research design and Methods.

"Educational research can and does make a difference, but it will succeed only if we recognise its messy, contingent, fragile nature. Some policymakers believe that supporting educational research is crazy, but surely the real madness is to carry on what we have been doing, and yet to expect different outcomes.". (Black &Williams, 2003, p 635)

2.1 Research Design and Methods

2.1.1 Aims of the thesis

The main aim of the thesis is to evaluate the utility of using health economic methods in education. This is based on the rationale that education provision needs to be evidence informed and the growing need for schools and policymakers to demonstrate value for money. There is a need for more economic evaluations to accompany effectiveness research to support decision makers in education (Hummel-Rossi & Ashdown, 2002). In this thesis, the application of economic analyses concentrates on the evaluation of a regional Formative Assessment Implementation Project (FAIP) in North Wales. The purpose of this chapter is to describe the FAIP project, and the chapter details the rationale behind the chosen research design and methodology for the empirical studies in Chapters 4 and 5. The methods for the systematic review will be discussed in detail in the following Chapter 3.

2.1.2 Research Questions

This thesis is focused on the following research questions:

RQ1: What is the evidence for the effective use of health economics approaches to evaluate education programmes/provision?

RQ 2: What is the impact of the Formative Assessment Implementation Project (FAIP), and does this represent value for money?

RQ 3: What is the feasibility of using health economics approaches to evaluate a large-scale education programme in schools?

Table 2.1 details an overview of the research questions, and the research design and methods used in this thesis. It also shows how the individual studies in this thesis (Study 1 to Study 4) relate to the main research questions. Study 1 aims to evaluate the extent to which

health economics methods and concepts have been used in an education setting and addresses research question one. The purpose of the primary research was to address research questions 2 and 3 to evaluate the FAIP and demonstrate the use of a health economics methodology in an education setting.

Table 2. 1 Overview of research questions design and methods.

Secondary research – Systematic Review (Study 1)	
Research question	**RQ1: What is the evidence for the effective use of health economics approaches to evaluate education programmes/provision?**
Research design: A systematic review methodology was used to answer RQ1. **Methods**: The systematic review was registered with PROSPERO (Reg: CRD42019123564). Reporting follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2015), and utilised the Drummond et al. (2015) 10-item quality appraisal checklist was used for the included studies. A total of 13,704 studies were reviewed at the title and abstract stage. A total of 12 studies were included, and a descriptive analysis was employed.	
Primary research-Interviews (Study 2)	
Research question	**RQ 2: What is the impact of the Formative Assessment Implementation Project (FAIP), and does this represent value for money?**
Sub- questions	What are the perceptions and experiences of Tier two teachers in relation to implementing formative assessment strategies and the perceived impacts in the classroom? What are the barriers and facilitators of implementing formative assessment in the classroom?
Research design: To evaluate the effectiveness of the FAIP, a Theory of Change methodology was employed (Chen, 2006). **Method:** Interviews with teachers were used to gain an insight to their perceptions of implementing FAIP and the impacts. Semi-structured interviews were conducted with seven teachers and analysis was undertaken using a Thematic Analysis approach (Braun & Clarke, 2012).	
Primary research-Focus groups (Study 3)	

Research question	**RQ 2: What is the impact of the Formative Assessment Implementation Project (FAIP), and does this represent value for money?**
Sub- questions	What are pupils' experiences and perceptions of using a range of formative assessment strategies?

Research design: The research design behind Study 3 is the same as Study 2.

Method: Focus groups were used to gain the perceptions of the pupils in the FAIP and to investigate their understanding of the different strategies. They were analysed using thematic analysis (Braun & Clarke, 2012).

Primary research-Cost Consequence Analysis (Study 4)	
Research question	**RQ 3: What is the feasibility of using health economics approaches to evaluate a large-scale education programme in schools?**

Research design: Study 4 used a quasi-experimental design to evaluate the impact over time. The measures for the quasi-experimental design were incorporated into the Cost Consequence Analysis. For the Cost Consequence Analysis, the perspective was from the education provider. Cost data was supplied from the education provider who were disseminating the FAIP project.

Methods: Pre-post **quantitative data** was collected from intervention and control schools to compare the impact of the FAIP on learner outcomes. A 2 x 2 mixed-model ANOVA was conducted, and mean scores are reported for all quantitative measures. No imputation was conducted on missing data. **Observational data** was collected from classrooms in the intervention arm. The **qualitative findings** were also included in the CCA to give the decision maker all the impacts from the intervention.

2.2 Evaluative framework for Studies 2, 3 and 4

Ethics

Ethical approval was granted for all studies from Bangor University Psychology Ethics and Research Committee (application number: 2018-16324-A14505).

2.2.1 Theory of Change

Given the scale and dissemination of FAIP, this thesis employed a theory of change model to conceptualise and describe how/why the desired change was expected to happen in the context of the FAIP project.

Theory of change works particularly well for community initiatives given the difficulty to evaluate initiatives in real world settings. Given the complex mix of different strands that the project will face (political, economic, and social) and will operate on different levels (institutional, school network and individual) that can interact with the project. Theory of change involves working with stakeholders to explore what they perceive to be the intended outcomes (Chen, 2006) and to discuss whether the project will in the first instance meet the following aims:

- "It should be plausible. Do evidence and common sense suggest that the activities, if implemented, will lead to desired outcomes?
- It should be doable. Will the economic, technical, political, institutional, and human resources be available to carry out the initiative?
- It should be testable. Is the theory of change specific and complete enough for an evaluator to track its progress in credible and useful ways?" (Connell & Kubisch, 1998, p. 3)

Using a theory of change supported the identification of time scales, intended outcomes, the necessary inputs, and suitable methods to understand the impact or possible outcomes. One way to articulate the theory of change is in a logic model, a logic model has been used to support policy implementation throughout the world (Auriacombe, 2011) and has been used in the Scottish Education system to articulate the recovery from the pandemic and implementation of their attainment challenge policy (Education Scotland, 2022). Through consultation with GwE, a logic model was developed (see Figure 2.2) to clarify the stakeholder's intended outcomes.

Figure 2.1 1: Theory of Change (Logic model)

Context	Aims	Rationale/theory of action
• Wales is undergoing a radical policy change with a new curriculum being implemented over three years (Donaldson, 2015). • Wales had come under international scrutiny with regards to the outcome on the Program for International Student Assessment (PISA). • Initial teach training, had been reviewed and there is a movement to get evidence informed practice into schools (Furlong, 2015). • National priorities require Welsh schools to reduce the attainment gap using more evidence - informed strategies and ideas while having a stronger emphasis on formative assessment (Welsh Government, 2017) . • Evidence indicates that formative assessment strategies have the potential be very promising approach to help improve standards of teaching and learning and to reduce the attainment gap, particularly for pupils eligible for free school meals eFSM (Williams and Black,1998)	• Improve the quality of teaching standards region wide. • Reduce the variance in teaching. • Embed effective formative assessment strategies in schools. • The development of pedagogy through reflection and being open to challenge,. • To have a consistent approach to formative assessment approaches across the region. • Raise standards in line with national priorities. • Reduce teachers workloads in response to national agenda. • Prepare for the curriculum, the 4 purposed and the 12 pedological principles. • Respond the 5 dimensions of professional standards for teaching and leadership(Welsh Government, 2017)	• Lead teachers attending the training in principles of effective formative assessment. • Lead teachers trial formative assessment strategies within their classrooms and reflect on the impact, adjusting and/or improving strategies as required. • Lead teachers will disseminate strategies throughout the school and to partner schools in Tier 2 and 3. • The quality of teaching and learning will improve. • A reduction in teacher workload as more effective feedback strategies are employed.

Assumptions
• All lead teachers will attend training and apply the principles in their classroom. schools will release teachers for the duration of the project. • Teachers will have capacity and support within school to implement strategies and be supported by school leadership teams. • Support from GwE and lead schools to advise Tier 2 and 3 schools as the program rolls out. • Consent will be obtained to track and monitor pupil attainment, as well as administer questionnaires.

Inputs	Activities	Outputs	Participants	Outcomes/impact
• KESS studentship • Supervisory team • Research student • Continued finical support from GwE • School engagement • Teacher engagement • Parental consent to collect data • GwE staff to carry out observations	• Bangor University to grant ethical approval • Consent from teachers and pupils to take part in the research project • Teachers to attend GwE training • Tier one teacher to provide support for tier 2 schools • Teacher can trial strategies within their schools • Collect pre/post data from pupils • Collect attainment data from GwE • Statistical support for PhD	• Teachers will be trained to used effective formative assessment methods. • Analysis of CHU-9D, SDQ, QoLS data in relation to effectiveness. • Interviews and focus group data analyzed to look at impact. • Attainment data to be analyzed • Impact report submitted to GwE.	• Teachers to trial and develop formative assessment strategies. • GwE to continue supporting schools • Tier one teachers provide support to Tier 2 schools. • PhD researcher and supervisory team. • Pupils with parental consent to participate in questionnaires and focus groups.	**Short :** • More effective use of formative assessment strategies in schools. • More effective teaching leading to improved engagement and levels of participation by learners. **Medium :** • More effective teaching leading to improved understanding of concepts and ideas. • Improved outcomes. **Long :** • A sustained improvement in the quality of teaching and embedded formative assessment principles in schools. • Sustained improvement in outcomes over time, with more learners achieving their potential.

2.4 Research design for Studies 2 and 3
2.4.1 Qualitative methods

Qualitative methods were used to evaluate the impact of the FAIP for RQ 2 and RQ 3. Interviews were conducted with teachers for **Study 2, and 3.** The results are also discussed briefly in **Study 4** to help conceptualise the entire evaluative framework and support decision makers.

2.4.2 What is qualitative research

Malterud (2001) describes qualitative research as a systematic approach to collect data mainly from talk or text. Qualitative research explores individual or group meanings around a social phenomenon that the participance have experienced. Qualitative research develops a picture of the experiences and perceptions of participants on a given topic uncovering a depth and richness to data that quantitative methods cannot. While there are criticisms faced at qualitative research including the lack of robust data collection methods, particularly at the validity and reliability, qualitative research has developed robust ways to address the criticisms, inter coder reliability, member checking and being reflexive in the research process supports the validity and reliability of findings (Grossoehme, 2014). Qualitative methods can support the research to investigate the social validity of an intervention as not all behaviours are quantifiable or observable (Braun & Clarke, 2012). Social validity is the acceptability, importance, and validity of an intervention, and moves from its origins in applied behaviour analysis to move beyond 'clinical judgments' (Kennedy, 1992, p. 147) to a broader focus on the social environment of the individual.

2.4.3 Semi structured interviews: Study 2

Semi structured interviews were used to gather the perceptions and experiences of teachers about the FAIP (Kallio et al., 2016). A semi structured interview can be defined as an exchange that takes place with a participant and a researcher. Open ended questions are used to understand and investigates the participants experiences and perceptions around a defined subject (Adams, 2015). While other methods are available for example unstructured interviews which are focused on eliciting social realities from participants without any pre-defined concepts often used in anthropology and sociology (Zhang & Wildemuth, 2009). Given that the aim of this research was to understand the perceptions held around formative

assessment, semi-structured interviews allowed this while still affording participants the space and time to discuss elements that were important to them (Adams, 2015) as well as check for clarity in the participants responses. Participants were also asked about time and cost in the interviews as not to burden them with a questionnaire that would take further time away from teaching.

2.4.4 Focus groups: Study 3

Study 3 consisted of focus group interviews with the pupils from each of the intervention classes and was considered an efficient method of collecting data in terms of time and cost (Kid & Parshall, 2000). Focus groups can be defined as a planned discussion with more than one member that is designed to elicit experiences or perceptions in a safe and non-restrictive environment. The researcher acts as a moderator of the focus group and facilitates the discussion using some pre-defined topics (Gibson, 2007). Moderators need to be able to lead the focus group and have the skill to be able to make the participants feel at ease. Extra consideration is needed when researching with children, the research needs to understand the unequal power balance that could lead to children answering in socially desirable ways. This is also true with children not wanting to object to peers' opinions to create favourable group responses, a moderator must have the skill and experience to be able to overcome some of these issues when researching with children (Punch, 2002). The decision to include the pupils was based on there being limited research on the perspective of pupils in relation to formative assessment research (Cowie et al., 2018). Also, as pupils were to benefit from the research it is also important to give the pupils a voice in the evaluation framework, and pupils' voice in research can provide powerful perspectives and generate knowledge around the given topic. Given pupil voice in the research addresses the unequal power balance that adults have in the design, implementation, and results (Lewis, 2007; Jones & Stanley, 2008).

2.4.5 Thematic analysis

For Studies 2 and 3, a thematic analysis approach was used to evaluate the information gathered. Thematic analysis is a technique to analyse and develop themes using qualitative data (Vaismoradi et al., 2016). This comprises of a systematic way to apply codes, exploring participants meanings with the goal of understanding and describing social reality of the topic under investigation. The themes generated from the data set allow researchers to understand the perceptions and experiences of groups of participants, and the themes are

developed across the whole sample rather than each individual participant (Braun & Clarke, 2012). In this study an inductive thematic analysis approach was used that allowed themes to emerge from the data, rather than be constrained by theories or hypothesis. The development of themes should also have the ultimate focus on answering the research question, and the research needs to be mindful of this in the coding and theme development. It is important that the research process demonstrates reflexivity in the data interpretation process (Byrne, 2022).

Braun and Clarke (2012) identify six stages to thematic analysis and is widely used in research outside of the original focus of sports exercise. First, the researcher needs to familiarise themselves with the data, either through transcribing the data oneself or reading transcribed data or listening to the audio of the dataset several times for each participant. Researchers may at this point notice some trends emerging from the data and could possibly note them down (Byrne, 2022). Second, following the identification of codes on each data set, the researcher labels relevant sections of text, and notes any information that is relevant to the research question.

Stage three is the generation of themes from the coded information, the focus is on aggregated meanings from the data set across participants. The themes need to be distinct enough from each other and start to build a picture of the commonalities across the data set. The researcher might at this point produce a thematic map of each participant, this is a fluid process, and the research may move the themes, combine them, or infer that their theme lack sufficient backing with little data to support it (Braun & Clarke, 2012). Stage four consists of reviewing the themes, this can be done at two levels, participants themes and themes across the data set, by reviewing the codes. Again, at this point there may be codes or data that does not support the theme and revision can take place. When this stage is finished the researcher will have a clear indication to the themes that are able to tell the story from the data and how the research questions are supported in the dataset.

Stage five is the final development of the themes, articulating the themes with the relevant extracts of the data set. The narrative of the data set is developed with a focus on answering the research questions, there themes should be concise and articulate the meaning of the theme to the reader. The final stage is to produce a report or the write up stage, there is a considerable overlap between stage 5 and 6. Even at this stage there can be revision but as with other stages the changes need to be documented to show the fluid and reflexive approach needed to produce qualitative finding. The researcher may use quotes from participants and should support the narrative as well as the overarching research questions. This research was part of a studentship and supervisors were able to offer support with theme

development as well as making sure that the research questions were addressed. Developing themes and codes for data is a fluid approach, appendix A1 demonstrated the iterative process the research went through and the reflective process.

2.5 Research design for Study 4

2.5.1 Quasi-experimental design

A quasi-experimental pre-post design was used with matched control schools as a comparator. Given that the researcher was not able to randomise the study into the intervention and control group as the intervention was already being disseminated. A quasi-experimental design is appropriate when programme delivery is underway independent of the research project (Campbell & Stanley, 2015). This type of research design has been used extensively in large scale evaluation studies and for policy implementation studies (Hakim, 2000; Handley et al., 2018). These designs are also used in public health where there is less control over the participants receiving the intervention. Gray (2004) suggests that quasi experimental designs work well in an educational setting as the researcher has no/limited control over the formulations of a classes.

2.5.2 Mixed methods

Mixed methods in its simplest form are a combination of qualitative and quantitative data collection methods. The qualitative and quantitative methods complement each other, particularly when used with a complex intervention (Petticrew et al., 2013). Using mixed methods can maintain some validity, identify implementation issues, and explain complex interventions (Doyle & Byrne, 2016; Zohrabi, 2013).

2.5.3 What is quantitative research

Quantitative research centres on generating knowledge and testing hypothesis mainly based on collecting data that can be converted to numerical values. Quantitative research uses data to measure a specified population. To integrate the numerical data the research uses in most cases statistical techniques to, explore, interpret, and present data (Sukamolson, 2007). For instance, quantitative data can be collected from questionnaires or surveys, around a given topic and the research (usually with statistic software) analysis, interprets and presents findings. Quantitative research is concerned with the generalizability, replication and

causality from the data collected (Bryman, 2016). The main criticism of quantitative research comes from an epistemological and ontological standpoint where concerns around objectivity decontextualises individuals from the 'real world' (Bryman, 2016).

2.5.4 Research measures: Study 4

As mentioned above a quasi-experimental design was used as the FAIP intervention had already started. The research focus on the second tier of the implementation. Given previous research quantitative measures were employed to understand if any impact could be observed using questionaries and national test data. Qualitative research will support the interrogation of costs and outcomes in an economic evaluation.

Ethics

Ethical approval was granted for all studies from Bangor University Psychology Ethics and Research Committee (application number: 2018-16324-A14505).

2.5.5 Measures

The Child Health Utility-9D

The Child Health Utility-9D (CHU-9D) was developed for children aged 7-17. This questionnaire measures the health-related quality of life and has been validated as a preference weights measure for children (Stevens, 2012). CHU-9D covers nine domains: worried, sad, pain, tired, annoyed, schoolwork/homework, daily routine, and the ability to join activities, with 5 levels of response (Boyer et al., 2014). Instruments like CHU-9D can be used to calculate utility weights, and these are commonly used in healthcare evaluations. The utility generated is used to calculate a Quality-adjusted life-years (QALY). The QALY can capture quality of life and length of life from interventions and is commonly used in healthcare cost-effectiveness and cost utility research (Furber & Segal, 2015). The questionnaire is available in multiple languages allowing straightforward dissemination in both Welsh and English so that the pupils could choose what language they wanted. A copy is in appendix A2.

The Strengths and Difficulties Questionnaire

The Strengths and Difficulties Questionnaire (SDQ) is a child and adolescents screening tool commonly used to identify behaviour difficulties. SDQ can be used with

children aged 4 to16 years, together with a modified version for younger children (aged 3 to 4 years). There is version for parental or teacher/carers to complete as well as self-completion versions (White et al., 2013). Self-completion has been validated for ages 8 and above (Muris et al., 2004). There are 25 questions covering five domains; emotional, conduct problems, hyperactivity, peer problems and pro social behaviour. The SDQ has been used extensively to assess social and emotional wellbeing as well as behavioural difficulties (White et al., 2013) the questionnaire is available in both Welsh and English from the developers. Formative assessment has been linked to improved pupils' behaviour, self-regulation, and non-cognitive improvements (Black & Wiliam, 2009; Dunphy, 2010). A copy is shown in appendix A3.

The 'Quality of Life in School Questionnaire'

The 'Quality of Life in School Questionnaire' (QoLS) was developed originally in Israel by Weintraub and Bar-Haim Erez (2009) and covers four main domains; teacher student relationship and social activities (12 items); physical environment (11 items); negative feeling towards school (eight items); and positive feelings towards school (five items). The QoLS questionnaire has four level response rating and is primarily aimed at measuring pupils' wellbeing and satisfaction at school (Ghotra et al., 2016). This questionnaire has been translated into American English. Through consultation with the developers, revisions were made (see appendix A4 for revised language). At the request of the developers, the revised version of the questionnaire was translated into Welsh and then reverse translated back to English to check for accuracy. Some evidence suggests that pupils develop improved relationships with their teachers, enjoy school and feel connected (Ferreira et al., 2014). A copy is shown in appendix A5.

National attainment data

In Wales it is a statutory requirement for pupils in Years 2 to Year 9 to take National Reading and Numeracy tests. Pupils are tested on reading comprehension in either Welsh or English depending on their language of instruction. There are two numeracy tests that are administered to pupils: the Numeracy (Procedural) test focuses on measuring data and number skills; and, the Numeracy (Reasoning) test focus on pupils' ability to complete problem-solving tasks (Welsh Government, 2019a). The results of the test are presented as an 'age standardised score' (that takes the average score across all learners of the same age in years and months who take the test that year) and a progress measure which compares a pupil to all the pupils taking the test in that year in their year group. The results from these tests can

be used to identify pupils who may need additional support. Formative assessment has been linked to improved outcomes for pupils so national attainment data was collected. (Black & Wiliam, 2009; Clarke, 2014; Ozan & Kincal, 2018).

2.5.6 Analysis

Data was analysed using SPSS 25 using an Analysis or Variance (ANOVA) approach. An ANOVA compares the means scores collected from the different conditions in the research study. A 2 x 2 mixed-model ANOVA was conducted on all measures (Pallant & Manual, 2011).

2.5.7 Effect size

The survey measures were analysed and presented as mean scores (average) and standard deviation (*SD*). This will allow the identification of changes between pre- and post- outcomes for the intervention and control group. A *p*-value is presented alongside the mean and *SD*; with the *p*-value indicating the probability of the null hypothesis.

2.5.8 Missing data

A complete case analysis was conducted this means excluding responses where participants had missing responses (Jakobsen et al., 2017). The benefit of only using complete case analysis is to minimise bias. In particular, the developers of the CHU-9D discourage imputation of missing data (Harrington et al., 2010). During the administration of the survey, participants were given the option to omit answers. Missing data is common in education settings, and some studies will impute the missing data with imputation models (Fleming, 2011). In total 638 pupils took part in the surveys, 59% of the cases were omitted from the analysis due to incomplete post test data.

2.5.9 Observations

Using observational data can support substantiating the findings from other methods of data collection (Zohrabi, 2013). The use of observational data supports the triangulation of the different data points (Fielding, 2012). To identify if the teachers were delivering elements of the training program, challenge advisors from GwE were asked to carry out observations to identify the key elements of formative assessment (as described in Table 2.2).

We used a standardized observation checklists adapted from a checklist devised by Leahy and Wiliam (2014). GwE school improvement officers were trained on how to use the

observation checklist. The training consisted of a 30-minute session with the GwE school improvement officers presenting the form, explaining the purpose of the checklist. allowing them to orientate themselves with the form and then ask questions (see appendix A 6 for observation checklist). The checklist was also translated into Welsh. The observations checklist was focused on the following key principles:

Formative assessment checklist questions and areas of focus

- Is it clear what the teacher intends the students to learn?
- Does the teacher identify student learning needs?
- Do students understand what criteria will make their work successful?
- Are students chosen at random to answer questions?
- Does the teacher ask questions that make students think?
- Does the teacher give students time to think after asking a question?
- Does the teacher allow time for students to elaborate their responses?
- Is a whole-class response system used?
- Is teaching adjusted after gathering feedback from pupils (data collection)?
- Is there more student talk than teacher talk?
- Are most students involved in answering questions?
- Are students supporting each other's learning?
- Is there evidence that various forms of teacher feedback advance student learning?
- Do students take responsibility for their own learning?
- Does the teacher provide oral formative feedback?
- Does the teacher find out what the students have learned before they leave the room?

For each of the statements above, there were five levels of response as follows:

- Not applicable,
- Applicable but not observed,
- Observed but poorly implemented,
- Observed, and reasonably implemented,
- Observed and well implemented.

2.6 Cost analysis methodology: Study 4

2.6.1 Cost-consequence analysis

Cost-consequence analysis (CCA) lists costs and outcomes in a disaggregated way so that decision makers can choose the outcomes that is the most important to them. CCA allows the decision maker to form their own opinion and prioritise which costs, and outcomes, are important to their context (Charles & Edwards, 2016).

2.6.2 Rational for CCA

CCA is recommended particularly for public health and non- health settings as the perspective can be broad and there is flexibility in demonstrating a range of outcomes (Charles et al., 2019; Haghparast-Bidgoli et al., 2021; Hartfiel & Edwards, 2019). CCA has the flexibility to demonstrate a range of outcomes alongside costs and gives the decision maker the choice to consider the outcomes that are most relevant to the context or objectives. Coast (2004) suggests that CCA is a more useful form of economic evaluation particularly for social decision making, and the intervention does not need to be reduced to one outcome, unlike CEA (Charles et al., 2019). CCA is often suited for non-technical audiences and provides a clear way to present costs and outcomes in a disaggregated way (Hartfiel & Edwards, 2019). There is a difficulty in evaluating complex interventions that are delivered at multi-sites, where the intervention has different components and subject to change depending on the needs or environment of the setting. Within complex interventions there is also the challenge of quantifying benefits/outcomes into monetary values, omitting benefits/outcomes due to the difficulty in calculating monetary risks not supplying the decision maker with the important elements of the intervention (Rogers et al., 2008). Hummel-Rossi and Ashdown (2002) suggested that the 'qualitative residual', or the additional benefits that are difficult to quantify in financial terms, should be included where possible to support decision makers within educational research.

Given the complex intervention that is being evaluated in this study, a CCA was utilised to demonstrate the cost and outcomes (and given the difficulty placing monetary values on some outcomes in this study). Using CCA will allow the decision makers to decide what outcomes are relevant to their context or objectives. The FAIP intervention can be considered complex given the nature of the training and the fact that participants were able to adapt the intervention to suit their school context. Each school setting is represented by different staff experiences and abilities, ethos, structures, and environment, and this adds an additional layer of complexity onto the challenge of delivering the FIAP training across a complex, regional setting (Glouberman & Zimmerman 2002; Rodgers et al., 2008).

Cost collecting methodology

Cost collecting methodology in health falls into two broad categories; one is 'Top down' or 'Bottom up'. Top-down costing is where retrospective costing is used on secondary data to calculate costs sometimes called relative value units (RVU) although a helpful cost method, it can limit the precision of costing (Chapko et al., 2009). Bottom-up costing is calculating all the cost associated with running an intervention or treatment and can be advantageous for the transferability to different sites or treatment pathways (Morris et al., 2012) sometimes termed activity-based costing (ABC). Different ways of costing produce very different results in the cost outcomes, it is the job of the analyst to determine the correct cost collecting methodology, for example if the analysis is not concerned with variation of local cost, then top down would be suited, but if the variation on a local level is a consideration, then bottom up may be suited (Chapko et al., 2009). Collating costs was from an education provider perspective (GwE). The majority of costs were incurred by GwE, and the main cost was supplying costs to cover teachers to attend the training and for staff cost to support the project.

Collating costs

To collate the cost of teacher's time, the business as usual (BAU), budgetary data was collected from three local authorities in North Wales (Anglesey, Conwy, and Wrexham), budgetary data on maintenance costs and school services was utilised, this gave the cost for a teacher to stand in a maintained safe classroom 100% of the time with the necessary equipment. Salary information was gained from the teaching union in Wales on the average salaries for qualified teachers without senior leadership responsibilities, or additional teaching and learning responsibilities a mean was calculated from the main ranges (£27,018 - £41,604 or scale M1- Max-U3). Further on-costs were used from a previous study (Harden, 2019) to allow for the cost of pension and NI contributions. In Wales, teachers are contracted to work for 1265 hour a year, so the yearly total was divided to calculate the cost per hour. Detailed costing is provided in appendix A7. Table 2.2 details the calculation for costs of teachers'(BAU) time. Other researchers in education have faced the problem of collecting BAU costings (Scammacca et al., 2020). Given that in education there is limited established cost, a pragmatic approach was taken to calculate the cost of BAU, of teachers' time.

Table 2. 2 Cost for teachers' time including on costs.

Cost (Mean)	2018-2019 prices	2020- 2021 prices	2022-23 prices
Teacher cost yearly	£58,544	£60,947	£72,233
Cost per pupil yearly	£3183	£3314	£3928
Cost per hour	£46	£47.89	£56.89

Sensitivity analysis

To test assumptions, sensitivity analysis was conducted on three different parameters of the costing assumptions. Testing assumptions can support robust conclusions (Levin et al, 2017). The costs were supplied from GwE and are fixed cost and represent the actual cost, as with the other resources such as, translation for training and hiring training venues. Sensitivity analysis was conducted on the average pupils exposed to the intervention, the out-of-pocket expenses (OOP), and the cost of buying out teachers to attend the training and showcases using the BAU cost in the table above Table 2.2.

2.7 Intervention: Formative Assessment Implementation Project (FAIP).

From 2017-20, GwE delivered a regional formative assessment project across six local authority areas, incorporating primary, secondary, and special schools. The project was designed and delivered by GwE alongside a expert trainer and focused on the application of effective formative assessment strategies in schools. The studies in this thesis were developed to evaluate the second tier, or phase, of the FAIP project.

2.7.1 What is Formative Assessment:

The central principle of formative assessment is the use of feedback from pupils' learning to enable teachers to adjust instruction to ensure they optimise learning outcomes (Wiliam, 2020). The ultimate focus of formative assessment is to inform and improve teaching and learning' (James, 2017, p. 3) and improve the standards that learners achieve. According to James (2017) effective formative assessment consists of three stages:

Making observations: The teacher needs to explore what the pupil does or does not know, and this is typically achieved by listening to pupils' responses, observing pupils on task, and/or assessing class or homework tasks.

Interpretation: The teacher interprets the skill, knowledge, or attitudes of the pupils.

Judgement: Once evidence has been gathered through observation and interpretation; the teacher then makes a judgment on the next course of action to move the pupil forward.

While the three stages are core features of formative assessment, it is important to clarify how they can be integrated into practice. Bennett (2011) and Leahy and William (2015) describe five key elements for the effective translation of formative assessment principles into classroom practice:

1. **Sharing Learning Expectations**: What the pupil is going to learn and the success criteria to achieve the learning.
2. **Questioning:** Effective questioning to facilitate learning (not recall).
3. **Feedback**: Provide feedback that enhances learning, within the moment.
4. **Self-assessment:** Allowing pupils to take ownership of their learning, reflecting their own thinking.
5. **Peer assessment:** Facilitate opportunities for pupils to discuss their own work for the purpose of enhancing, understanding and knowledge.

2.7.2 Intervention

The formative assessment implementation project (FAIP) was commissioned by GwE to embed formative assessment practices across schools in North Wales. An expert in formative assessment was employed with experience in teaching and delivering formative assessment training to schools, particularly through a train the trainer model. The expert trainer was commissioned to deliver training to teachers in the first cohort of schools (known as Tier 1, lead schools), then support subsequent phases of training and support sessions with two additional intakes of schools (known as Tier 2 and Tier 3 schools respectively). GwE oversaw the project alongside the expert trainer and offered support throughout to all teachers in the project. See figure 2.1.

Figure 2.1 2: Structure of FAIP.
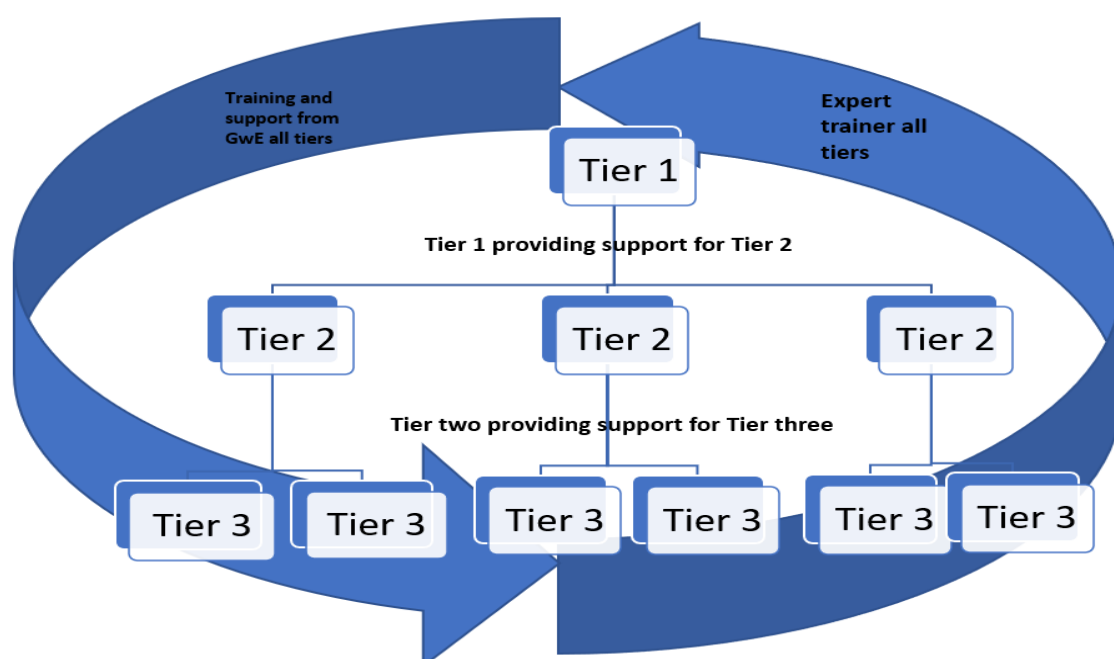
In October 2017 two teachers from each of the 27 Tier 1 schools attended the training and worked together to experiment the formative assessment strategies within their respective classrooms. One hundred and ninety-three Tier 2 schools received training in 2018-19 amounting to 386 teachers in total. And in the following year all the remaining schools in the region were invited to attend training and receive support from Tier 1 and Tier 2 schools 140 schools received training in Tier 3. The subsequent tiers were provided with support from Tier 1 teachers and supported the train the trainer model. Teachers were encouraged to share good practice, utilise support and guidance, and troubleshoot with the lead school. Tier 1 teachers presented to Tier 2 teachers at the training days and reflected on the strategies they had utilised within their class. While GwE facilitated and delivered the training, a core component of the programme was developing teachers' skills through a train-the-trainer model. Table 2.3 details the chronology of the project.

Table 2. 3Chronology of the project

Tier 1	2017 - 2018 academic year	**54** teachers from **27** schools initially selected through a process of application and interview. Training and collaboration led by GwE, and the expert trainer commenced school training in October 2017.

Tier 2	2018 – 19 academic years	**386** teachers from **193** schools collaborating under the guidance of Tier 1 schools and in conjunction with the expert trainer. Training and collaboration led by GwE, and the expert trainer commenced school training in September 2019.
Tier 3	2019-20 academic year	**261** teachers from **140** schools participated under the guidance of Tier 1 and 2 schools and in conjunction with the expert trainer. Training and collaboration led by GwE, and the expert trainer commenced school training in 2021-22.

The formative assessment strategies contained within the FAIP training session delivered by the expert trainer were drawn from a wide range of research-informed findings and teacher guides (Bennett, 2011; Clarke, 2014; Leahy & Wiliam, 2015). See table 2.3 (column1) for the core principles. The expert trainer presented 12 different strategies for the teachers to use, and the elements are described in Table 2.4 (column 2). Finally, the teachers were given practical ideas to help them implement the strategies in the classroom see Table 2.3 (column 3)

Table 2. 4 Summary of the formative assessment principles and practical ideas delivered in the FAIP training session

Formative assessment principles (Bennett, 2011; Clarke, 2014; Leary & Wiliam, 2015).	Practical classroom strategies shared by the expert trainer and GwE.	Strategies for teachers in the project to trial and research within their own classroom
Sharing Learning Expectations: What the pupils are going to learn and the success criteria to achieve the learning.	• Sharing the learning objectives with pupils, making sure they are clear. • Co-constructing success criteria. • Demonstrating examples of excellent work prior to pupils starting their own.	• Teachers to use out of context *learning objectives* to frame tasks. • Teacher to support pupils to co-construct skill-focused *success criteria*.
Questioning: Effective questioning to facilitate learning (not recall).	• Using effective questioning to elicit any prior knowledge or current understanding, this is especially important at the beginning of lessons.	• Prior Knowledge Questioning - to assess the learning needs, throughout the lesson with particular

	• Using peer and teacher feedback to show where excellence is, or improvements can be made. • Using the end of lessons to reflect and summarise on learning.	attention to the beginning of the lesson.
Feedback: Provide feedback that enhances learning, within the moment.	• To continually explore where pupils are in their learning and understanding and adjust feedback or the direction of the lesson. • Ensuring provision for cooperative feedback and mid-lesson stops are used as effective feedback opportunities. • Ensuring that effective marking strategies provide timely and useful next steps for learning.	• Display examples of good/ excellent work, to discuss with pupils. • Verbal and written feedback – throughout the lesson, mid-lesson learning stops, self-assessment, feedback and peer feedback, using visualiser to discuss successes and improvements, 1-1 conferencing, drafting. • Reduce cognitive load – effective lesson design to reduce the cognitive load.
Self-assessment: Allowing pupils to take ownership of their learning, reflecting their own thinking.	• Improving the culture of the classroom, so that teachers and pupils develop more of a growth mindset, improving pupil's self-belief that they can succeed; and improving pupils' meta cognition skills. • Pupils being involved with planning to cultivate ownership and motivation.	• *Growth mindset* – Learning powers (meta cognition) - Learning zone diagram ('bullseye') • Eliminate *comparative rewards* – praise to be learning specific.
Peer assessment: Facilitate opportunities for pupils to discuss their own work for the purpose of enhancing understanding and knowledge.	• Creating talk partners, so pupils are resources for each other and so all pupils can be involved in class discussion and questioning. • Provide opportunities for pupils to select different levels of challenge within tasks. Setting high expectations for task outcomes and maintaining pupils' self-esteem.	• *Talk Partners* – to support peer assessment and discussions relating to classroom activities. • Eliminate *ability grouping*- arrange the classroom so that pupils can work with different ability pupils.

Training was delivered throughout the 2018-19 academic year. Three training days were held alongside two review sessions (Table 2.5). The review sessions were designed to share effective practice, discuss progress, provide next steps, and allowed teachers to receive feedback on their experiences. Finally, a series of showcases were organised so that the teachers could present the work they had undertaken in school. These showcase meetings were attended by the tier of schools starting the training in the next academic year, and teachers were expected to present a poster on one of the strategies that they had used in their class this gave the incoming teachers the chance to ask questions and share initial thoughts on joining the training. Training materials are in Appendix A8.

Table 2. 5Timeline for teachers in Tier two.

Academic year 2018-2019	
Training Day 1 with GwE and Tier 1 schools	September 2018
Presentation by Expert and a Showcase of Tier 1 teachers	November 2018
Revision Session 1 in the Tier 1 lead school	January 2019
Training Day 2 with GwE and Tier 1 schools	April 2019
Revision Session 2 in the Tier 1 lead school	June – July 2019
A series of 12 showcases across the region displaying Tier 2 schools' work.	October 2019

Chapter 3 Systematic review of economic evaluations in education settings

"Studies increasingly include back-of-the envelope estimates of per-participant costs, which serve to contextualize the return of an education intervention. More comprehensive cost-effectiveness analyses that account for both monetary and nonmonetary costs, such as the opportunity costs of educators' time, would go even further to provide policymakers with valuable information for making difficult decisions with limited resources." (Kraft, 2020, p. 246)

3.1 Summary

Unlike other fields of scarce funding, education researchers have yet to embrace economic analysis to its full potential to aid decision makers to make informed decisions. This systematic review explores what economic research has been carried out predominantly in primary and secondary schools maintained by state or government.

Until now, no systematic reviews have explored this topic. This paper aims to identify: (1) any economic evaluation methods that have been utilised in a school setting; and (2) how pupils' attainment is being measured. A total of 12 published full economic reviews were identified; eight of these were cost-effectiveness studies and the remaining four were cost-benefit studies. Measures used ranged from standardised tests, high school completion and researcher designed tests. Using the Drummond et al. (2015) checklist to examine the quality of these reviews, three were judged to be of poor quality, five were judged to be moderate quality, and the remaining four were judged to be good quality. The discussion highlights the technical issues including choice of comparators, cost analysis, and the presentation of findings that limit the scope to support decision makers. The chapter concludes by making recommendations for consistency when conducting and reporting of studies to support decision makers.

3.2 Introduction

For many years education systems in the United Kingdom have functioned in a challenging context due to pressure from increased accountability stemming from school inspections (Copland, 2019) and a framework of international comparisons set out by the Organisation for Economic Co-operation and Development (OECD) and Program for International Student Assessment (PISA; Brown et al., 2016). Coupled with decreasing financial resources, for example, Wales has seen in real terms spending decrease by 5% over the last ten years (Sibieta, 2020) Since the global economic downturn in 2008 there have been

significant austerity measures implemented by the United Kingdom government. Although funding for education was protected, increases in poverty and cuts to peripheral services have begun to take effect in education (Lupton & Thomson, 2015). Public services are challenged with 'doing more for less' and educators have faced the challenge of educating the next generation with increasingly scarce resources.

Some economists have evaluated the rates of return from a given education investment and have identified subsequent gains for the labour market in terms of higher skilled work force and additional tax revenue collected (Vella & Gregory, 1996). Concerns from the US government about the cost of a poorly educated workforce led to Levin and Buchman (1972) analysis of the financial consequences of pupils leaving school with low levels of qualifications, and the cost this had to society. The benefit of Early years education and/or pre-school programs have also gained economic attention in relation to their favourable rates of return (Barnett, 1985; Heckman, 2006). Such investments typically result in additional tax revenues, which can be redistributed back through government expenditure. The benefits generated; a more skilled workforce and enhanced employability, a reduction in crime and improved health outcomes amount to significant social benefits and a saving on public expenditure. However, given the large investments of public money, and the long-time span that these interventions need to generate impact, policymakers are often driven to seek quick fixes that operate over shorter timescales (see Harris [2009] for a detailed discussion on the issues of politics and education policy).

3.2.1 Specific issues within education

While top-down policy like the Pupil Premium aims to reduce the attainment gap, there is little guidance for how schools should spend the extra funding for disadvantaged pupils (Copeland, 2019). Although this extra funding is typically spent on extra staffing, there is no consensus on whether targeted or whole school approaches are more effective (Holton, 2017; Machin et al., 2013). In the case of both the PDG and the Pupil Premium Grant, schools were 'encouraged' to seek evidence-based interventions from organisations such as the Education Endowment Foundation (Hilton, 2017) and school leaders have considerable autonomy to identify and employ interventions that suit the needs of their pupils (Ahmed, 2020). Carrier (2017) highlighted that interventions that are promoted to school leaders are disproportionally based on anecdotal rather than empirical evidence. Gorard et al. (2020) highlighted that schools might struggle to elicit desired outcomes from effective strategies

due to the challenges of ensuring fidelity of implementation (leading to ineffective practice). Additionally, other researchers comment that education research lacks the tools to measure effectiveness appropriately or does not include a consideration of all the wider benefits of an intervention (Hummel-Rossi & Ashdown, 2002). School leaders and education policy makers face significant difficulties in identifying effective programs and/or interventions.

3.3 Review focus

There is currently very limited research that incorporates robust economic evaluations of education programs and interventions. This limits the ability of education leaders to make policy decisions based on information on effectiveness and cost. Clune (2002) carried out a review as to the breadth of cost-effectiveness analysis in education, although in the initial stages identified 1329 titles over a five-year period (1991-1996), and of the 541 titles selected for further review, 56 % were classed as rhetoric. Of the 541, none were classes as plausible and only four were classed as partly plausible[2]. Evaluating interventions in education without considering information on cost limits the quality of decision-making and prevents a more balanced judgement (Bowden et al., 2017; Hummel-Rossi & Ashdown, 2002; Levin, 2007). The palate for economic evaluations is hindered by the decision makers themselves, either that they are not politically attractive, or the evidence points to a strategy that is not welcomed by voters/parents (Clune, 2002; Levin, 2001). However, without information on intervention costs, decision makers are left to perform "back of envelope" calculations (Kraft, 2000). Levin (2001) hypotheses the lack of development of cost-effectiveness in education can be linked to the supply and demand of such evaluations. First, lack of training, second, the lack of effectiveness data (poor quality) and third, lack of demand from policy makers.

While economic evaluation has been used for many decades in the Army, other fields of public resource allocation have also adopted economic evaluation for decision making. An example would be that new treatments cannot be considered for the National Health Service in the UK without economic evaluation (Morris et al., 2012). Although some education researchers have looked at the economic methods used in health research and have suggested that they would be a useful addition to educational research (Hummel-Rossi & Ashdown, 2002), to date there has been no large-scale initiative to employ economic evaluation methods in education research. The field of health economics has a wealth of published research that

---

[2] Clune (2002) categorised the studies in the following way, Rhetoric: claims to be CEA but no data on cost or effects. Minimal attempt: contains minimal data on cost or effects. Substantial attempt: Contained cost and effect data but with serious flaws. Partly plausible: Either strengths in costs or effects. Plausible: comprehensive approach to resource use, effectiveness data was robust and used comparisons.

describes a range of robust methodologies for evaluating interventions, not only for effectiveness but also for cost-effectiveness. The purpose of this systematic review was to establish what economic methods have been used in school settings and how outcomes of pupil attainment are measured. Chapter 1 described the main methods of economic evaluations that are common in health care analysis: cost-effectiveness analysis; cost-utility analysis; cost-benefit analysis; cost-minimisation; and cost-consequence analysis. These methods have a shared characteristic of the measurement and valuation of costs, although they differ in their approach to the measurement and valuation of benefits.

3.3.1 Summary and main research question(s)

Review focus and inclusion criteria

Schools and policy makers need to use resources effectively to raise pupil attainment. To help them achieve this, researchers should support them to make informed decisions based on the evidence and analysis techniques available. Given the little attention that has been given to economic evaluation in educational research, and the call for economic evaluation in education, this review seeks to understand what type of economic evaluation methods in education have been reported in the research literature, including the breadth and quality of studies.

Aims

1. What economic evaluations have been undertaken school settings, and what methods were utilised, (e.g., CEA or CBA). And, whether these methods are used effectively based on a quality appraisal tool.

2. To identify outcome indicators used to assess educational attainment.

3.3.2 Methods

Selection of studies

Economic evaluations are evaluations that compare costs (resource use) and benefits (outcomes) with other courses of action. Studies were excluded that did not include a full economic evaluation, including those displaying cost but no comparator or incremental costs

and effects. Papers were reviewed where the interventions were aimed at school age pupils, within a school setting, and included educational outcomes.

Search strategy and identification of studies

To identify synonymous language in different countries (for example, school years 'Year 2/ Grade 1' or 'attainment/ achievement') it was important to establish appropriate search terms to avoid missing relevant studies. Appendix B1 provides details of the terminology used in this systematic review. The systematic review was registered with PROSPERO to reduce reporting bias and duplication (registration number: PROSPERO 2019, CRD42019123564, date of registration: 1 March 2019) see Appendix B2.

The following databases were used to search for papers: ProQuest Social Science Premium, Jstor, Web of Science, PsycINFO and Cochrane Library (including PubMed and Medline). Appendix B3 includes examples of the search strategies used for ProQuest. A broad search strategy was employed to obtain a comprehensive set of results. Papers were restricted to research conducted in OECD countries and in the English language, as these are more well-developed, comprehensive education systems that are more comparable. The sample was restricted to pupils aged 4-18 years. This reflects the typical ages that pupils, internationally, enter full time, compulsory education.

The search strategy consisted of terminology from the scoping searches, this gave six lists in relation to educational interventions, pupil educational outcomes, pupils, evaluations, schools and economic methods. Search terms for each list were combined with "OR", and each list were combined with "AND", terms were searched within title, abstracts and keywords. There was no restriction on the date studies were published.

The Journal of Education Economics and The Centre for Benefit Cost Analysis in Education were hand searched. Reporting follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2010). All returned searches were inputted into a reference manager. Once duplicates were removed, screening of titles and abstracts were completed by the first author (EJT), full texts were then accessed and independently reviewed by two reviewers (EJT and BFA), any discrepancies were resolved with the support of a third reviewer (JC). Two reviewers (EJT and JC) assessed the quality of the included studies using Drummond et al. (2015) checklist for economic evaluations see Table 3.2.

Quality Appraisal

A health perspective was taken for this review. Drummond et al. (2015) 10-item quality appraisal checklist was used for the included studies. This is a widely used and highly recommended quality appraisal tool for economic evaluations in health. Table 3.1 provides a summary of the quality appraisal for each included paper. A copy of the long version of the Drummond et al. (2015) 10-item quality appraisal checklist is available in Appendix B4. The PRISMA preferred reporting checklist was used to report the findings, this is detailed in Appendix B5.

Data extraction

Two reviewers (EJT and JC) developed the data extraction forms. The purpose of data extraction was to elicit the main characteristics of each included study (see Table 3.3).

Converting and inflating costs to 2023 prices.

Costs are converted and inflated to 2023 prices. For the conversion from dollars to pound sterling, OANDA was used. For papers older than 1990 conversions were sourced from Pound Sterling Live, a historical chart. Conversion calculations to inflate historical data to 2023 prices were made using The Bank of England inflation calculator. Cost are rounded to the nearest pound.

3.4 Results

A total of 24,179 studies were identified and transferred into a reference manager. Once duplicates had been removed, 13,704 studies were reviewed at the title and abstract level. Exclusion at this stage was mainly due to the study not associating costs to the program, authors stating that interventions were cost-effective but presented no analysis to support this claim, or lack of robust costings (only rough estimates) or no attainment measures. Twelve studies met the inclusion criteria for the review, and all were from America (8 were CEA; 4 CBA/BCA). One paper reported on both CEA and BCA, each method will be discussed separately in this paper. Figure 3.1 provides the flowchart of the search strategy.

Figure 3.1 1 Systematic Review flow diagram (Moher et al., 2010)

Identification

| Records identified through database searching (n = 24,179) | Additional records identified through other sources (n = 34) |

Screening

Records after duplicates removed (n =13,704)

Title and abstracts screened (n=13,704)

Records excluded (n = 13,530) Not economic evaluations

Eligibility

Full-text articles assessed for eligibility (n = 174)

Full-text articles excluded, with reasons (n = 162) Not full economic evaluations, no attainment measure, not an in-school intervention

Included

Studies included (n = 12)

Cost effective: (n=8) Cost benefit/Benefit cost: (n=3 Cost effective and Benefit cost: (n=1)
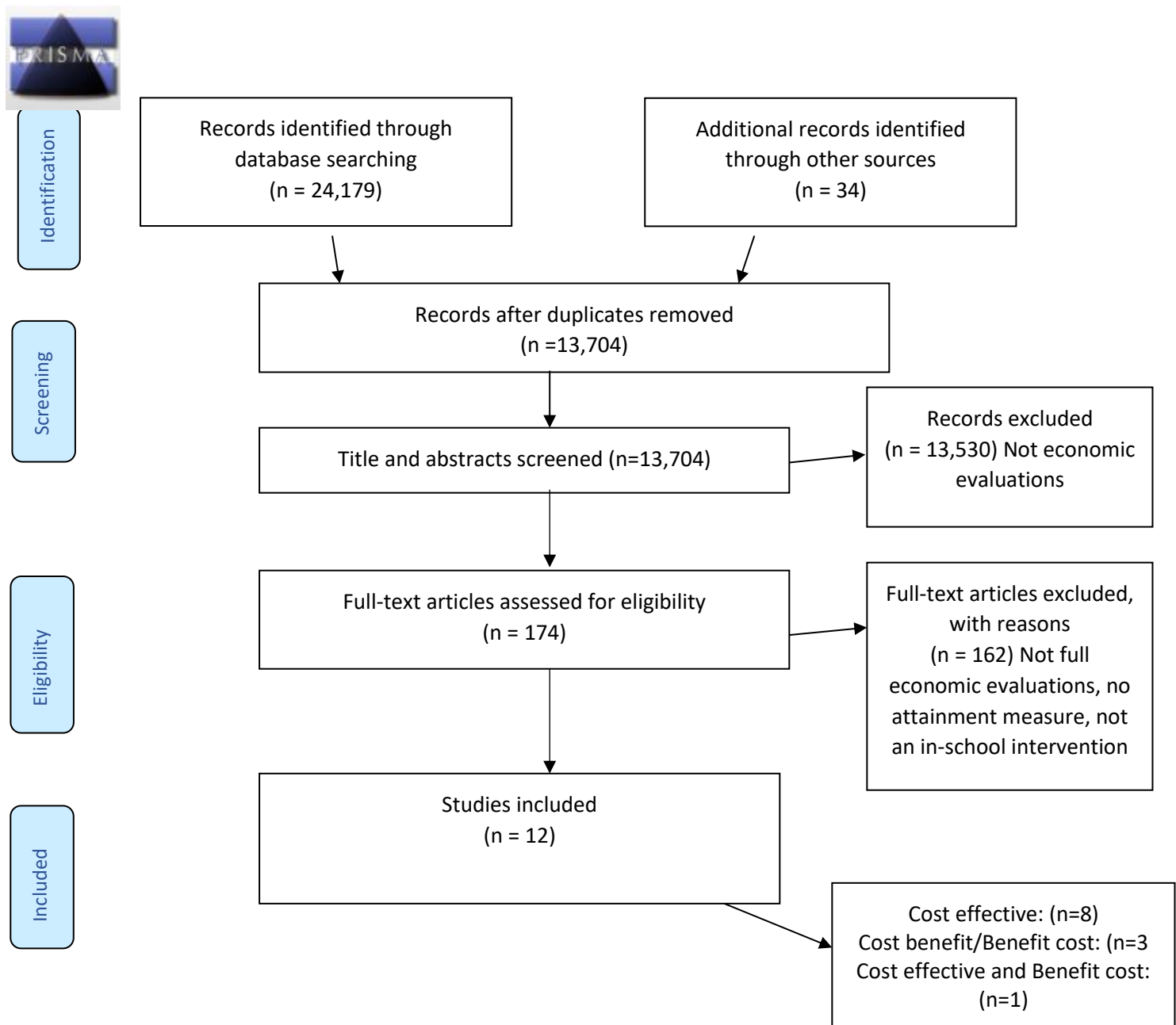
Table 3. 1 Quality Appraisal results.

The results of the quality insurance exercise are *n = 4* papers were of good quality scoring 9 +. Five studies were of a moderate quality, with a score between 6 and 8. The remaining papers (*n = 3*) were of poor quality with a score under 5.

Note: P = yes; O = no; - = can't tell, N/A = not applicable. Quality rating based on the number of Drummond questions answered: 0-5 = poor

Drummond questions	Yeh (2007)	Yeh, (2009a)	Belfield et al. (2015)	Quinn & Mondfrans (1984)	Reynolds et al. (2002)	Bowden et al. (2015)	Hollands et al. (2014)	Levin et al. (1987)	Hollands et al. (2016)	Yeh (2009b)	Borman and Hewes (2002)	Bowden & Belfield (2015)
Was a well-defined question posed in an answerable form?	P	-	P	P	P	P	-	P	P	P	P	P
Was a comprehensive description of the competing alternatives given?	P	N/A	N/A	-	N/A	-	P	P	P	P	-	N/A
Was the effectiveness of the programs or services identified?	P	P	P	P	P	P	P	P	P	P	P	P
Were all the important and relative costs and consequences identified?	O	P	P	P	P	P	P	O	P	P	P	P
Were costs and consequences measured accurately in appropriate physical units?	O	–	P	P	P	P	P	O	P	O	P	P
Were costs and consequences valued credibly?	O	O	P	P	P	O	-	-	P	O	P	P
Were costs and consequences adjusted for different timings?	-	P	P	O	P	P	P	O	P	-	P	P
Was an incremental analysis of costs and consequences of alternative performed?	O	N/A	N/A	O	N/A	O	O	O	O	O	O	N/A
Was allowance made for uncertainty in the establishment of cost and consequences?	P	P	P	O	P	P	P	O	P	O	O	P
Did the presentation and discussion of study results include all issues of concern?	O	O	P	P	P	P	P	-	P	-	P	P
Quality Score	4	6	10	6	10	7	7	3	9	4	7	10

quality, 6-8 = moderate quality, 9+ = good quality

3.4.1 Main findings

Cost-Effectiveness Analysis (CEA)

CEA was the most common economic evaluation method used (eight studies). One paper (Bowden & Belfield, 2015) conducted both CEA and BCA; for the purpose of this systematic review each method will be discussed separately.

Yeh (2007) conducted CEA of rapid assessment, a program designed to teach reading and mathematics, and provide rapid, frequent feedback to pupils. The program was compared to four other interventions: increased spending, charter schools, voucher programs, and accountability. Yeh (2007) collated retrospective data from previous evaluations of each program and sourced cost information from various research and government documents. Outcome measures for rapid assessment were program specific, the STAR Reading assessment and the STAR MATH assessments. The average effect size for rapid assessment was 0.32 standard deviations (*SD*) and the average cost per pupil, per year, was £23 including the opportunity costs. The cost for increased spending was £953 ($ 1,119) per pupil and an effect size 0.083 (*SD*), Voucher schemes costs were £9,213 ($9,646) per pupil with the effect size of 0.057 (*SD*). Charter schools' costs were £6,885 ($8,086) per pupil with the effect size of 0.005 (*SD*), and finally Accountability was £167 ($19) per pupil with the effect size of 0.050 (*SD*). The relative effectiveness-cost [3]ratio for rapid assessment was 0.014. Teacher time to monitor and the purchase of equipment were included in the sensitivity analysis, discount and interest rates were applied. Yeh (2007) acknowledges that rapid assessment is being compared with high cost and low effect interventions thus exaggerating the cost-effectiveness of rapid assessment.

Quinn, Mondfrans, and Worthen (1984) conducted CEA on two different mathematics programs, GEMS Math and Text Math. The purpose was to investigate which program was most effective for pupils with different levels of socioeconomic status (SES). Both programs had the same outcome measures: the Iowa Test of Basic Skills (ITBS) and the District Maths Test (DMT). The costs were calculated from interviews and questionnaires with district personnel and teachers. The results were complex; no one intervention was cost-effective for

---

[3] Relative effectiveness-cost ratio is an analysis that is employed by some authors in education, the effects are divided by the costs. The ratio provides the reader with how much £1 of effect is produced. This is then extrapolated up to demonstrate what can be 'purchased' for the extrapolated amount. This is not a strategy employed in health research.

all types of parental socioeconomic status (SES) students. The cost of each intervention per year was £1006 for GEMS Math, and £1,312 for Text Maths. The authors adjusted the scoring for guessing levels on the tests, this resulted in a negative effect for high SES pupils. Text Maths is more cost-effective for high SES pupil and GEMS maths is more cost effective for low SES. No sensitivity analyses were performed on the results and there were no adjustments made for costs or discount rates for timing. The authors conclude that they would have liked to measure additional outcomes in the form of pupils' ability, more comprehensive attainment data, and teacher morale.

Hollands et al. (2014) used the method of CEA to assess education interventions aiming to improve high school completion. They looked at five interventions that were targeted at supporting students to complete a high school diploma (HSD) or gain General Education Development (GED) as the outcome measure. The comparators were, Talent Search, Jobs Corp, JOBSTART, NGYC and New Chance. This was a retrospective study using results from previous research or US department of Education information. Cost information was obtained from other research or government available data. The research was conducted under a social perspective. Discounts and sensitivity analysis were conducted in the evaluation of Talent search. Talent Search is the most cost-effective of the dropout prevention programs with 3.3 extra graduates per £90,879 invested ($100,000), whereas the least cost-effective is New Start with 0.5 extra graduates per £90,879 ($100,000). The authors detail the issues of collecting relevant cost data and address the issue of comparability of Talent Search to the other interventions.

Levin, et al. (1987) used CEA to calculate the cost effectiveness of Computer Assisted Instruction (CAI) compared with cross age tutoring, class size reduction and increasing instruction time for raising pupil attainment. Effectiveness data was collected from other research conducted on CAI and costs were estimated by the authors where cost information was not available. Outcome measures for the comparators are not stated, only the effect size. The results were calculated at a effectiveness-cost ratio of 0.22 per £210 ($100) for Cross Age Tutoring and for CAI was 0.15 per £210 ($100). Cross age tutoring was the most cost-effective intervention, followed by CAI. There is no discussion of discount rates for loss of effect over time. Levin et al. (1987) suggests caution should be made in interpreting the results as this was only an evaluation of one assessment of CAI and other research studies could have different outcomes.

Hollands et al. (2016) demonstrated the method of CEA on two early reading interventions. This was a retrospective analysis and was completed on Corrective Reading

and the Wilson Reading System. The outcome measures collected were alphabetic, oral fluency and comprehension scores. The programs were identified for analysis based on criteria of similar population and outcome measures that were comparable. All resources were detailed from either previous research, program developers and publicly available information, internet searches or interviews with staff. Prices were sought for each resource at a national price to be able to compare the interventions; there is a detailed breakdown of where the costs were found and the discount rates. The results suggest that Wilson Reading System is more cost-effective than Corrective Reading, with a cost of £18,498 for 0.33 effect size gain for Wilson Reading Systems and £41,884 for effect gain of 0.22 for Corrective Reading. Detailed information on discount rates and program life was provided by the authors, and sensitivity analysis was conducted on the alphabetic outcomes. The authors suggest that cost data should be routinely collected, and that the method of cost-effectiveness should be used more widely in education to aid decision makers.

Yeh (2009a) employed the method of CEA to compare the cost-effectiveness of Class Size Reduction and Rapid Assessment. Effectiveness and costs were calculated from previous research. The outcome measures used for Reduced Class Size were standardised tests and a program specific STAR Reading and STAR MATH assessment for Rapid Assessment. Yeh presented 'effectiveness-cost ratios' for both interventions; for Class Size Reduction the highest effectiveness-cost ratio was 0.0001[4] with an effect size of 0.125 and for Rapid Assessment the ratio was 0.0113[5] with an effect size of 0.391 *SD*. The average cost per student was £32 for rapid assessment. A relative effectiveness-cost ratio was calculated meaning that rapid assessment would achieve the effect 124 times faster than Class Size Reduction for every dollar invested in the intervention. No discount rates or effects lost over time were reported. No sensitivity analysis was conducted on the assumptions of costs and effects.

Borman and Hewes (2002) conducted a long-term CEA using Success for All. This intervention was compared with three other large-scale comprehensive school interventions. Tennessee STAR, Perry Preschool and Abecedarian Project Pre-school. The datum was collated from the original participants, and the control group information was provided by the school district. They used three outcome measures: attainment (on standardised tests), grade completion or retention, and placement of students in special education. Comprehensive

---

[4] Rounded to 4th decimal point
[5] Rounded to 4th decimal point

information on program costs were provided. There was no sensitivity analysis performed on costs or outcomes. The results were presented in effect per $1000 (£1373 at 2023 prices), with an effect size of 0.90 *SD* for reading and 0.40 *SD* for mathematics. Thus, Success for All was deemed the most cost-effective for the desired outcomes by the authors. The authors reduced the cost of comparators by the effect that they deemed would have been on mathematics and reading; this was calculated using information from the program developers of the comparators. Tennessee STAR intervention costs were reduced by 67%, Perry School intervention by 50%, and the Abecedarian Preschool project by 40%. The authors discuss the issues with this type of policy evaluation and that the benefits could take a long time to mature, and that a CBA would be suited to this type of evaluation.

Finally, Bowden and Belfield (2015) conducted a CEA as part of the research that they carried out evaluating the Talent Search program. This is a program aimed at disadvantaged students that seeks to increase high school completion or postsecondary enrolment. They evaluated the cost-effectiveness across the different sites where Talent Search was implemented. The main outcome measure was college enrolment or postsecondary education. Costs were collated retrospectively using information from senior leaders across the different school sites, and from publicly available documents using the ingredients method. The study identified mixed results between collage sites results, with some being cost-effective and some not being cost-effective. The average cost for Talent Search was £40,803 for a high school completer and £31,354 for each student enrolled in postsecondary education. The results were pooled across all sites, and an effectiveness-cost ratio of 2.3 high school completers per $100,000 (£89,984) estimated. The results are also presented in a cost-effectiveness plane[6], layered with a ceiling ratio[7]. Most of the simulated plots (78% for high school graduation, and 85% for postsecondary enrolment) fall within the North East quadrant of the cost-effectiveness plane, indicating that the intervention is both more costly and more effective than the comparator. The authors comment that there is a 'modest' chance that the intervention is not cost-effective. They conclude that further research on the variation of cost at each site is needed and that they could measure a wider range of benefits.

---

[6] Cost-Effective Plane is a tool used visually to support decision makers. A cost-effectiveness plane displays two or more treatments under investigation. The graph is sectioned into four quadrants and dependant on where the results fall, allows decision makers to assess easily if treatments should be accepted or rejected (Morris et al., 2012)

[7] Bowden and Belfield (2015a) used the cost benefit test as the ceiling ratio.

Cost-Benefit Analysis (CBA) and Benefit-Cost Analysis (BCA)[8]

Reynolds et al. (2002) conducted a CBA on the Chicago-Parent Centres. This is a comprehensive pre-and school age, intervention that targets families from disadvantaged areas. Outcomes measured were high school completion, potential earnings and tax revenue for the state. Reduction in crime and child abuse were also measured and costed into the findings. Savings to welfare were analysed as well as the extra cost of pupils entering college degree programs. This was a follow up study using data collated on the original participants for both intervention and control. Costing data was obtained from the original program costs and inflated to 1998-dollar prices. There is a good description of the sample, and the costs are extensive. The shadow pricing of benefits is extensive and appropriately sourced. There is also a cost breakdown for the pre-school and school age program. The results are presented as a benefit-to-cost ratio for each level of program and the difference in the sex of participant with a breakdown of where the saving fell. For the pre-school program there was a $7.14 (£9.56) return for a 1-dollar investment. For the school-age program, the saving per dollar invested was $1.66 (£2.38) and for the extended program it was $6.11 (£8.37) of the total benefit. There is a comprehensive sensitivity analysis and limitations discussed about the assumptions made in costing the benefits of the program. They suggest that the program is cost-effective and that although the benefits are not instantly tangible, policy makers can make use of the research for long term planning of provision.

Yeh (2009b) used CBA to demonstrate the economic benefit of investing in Rapid Assessment. The effectiveness and cost data were collated from his earlier work Yeh (2007). Yeh then used the effect size of 1.50 (*SD)* in relation to the Armed Forces Qualification Test (AFQT). An effect size of 1.50 *(SD) on* the AFQT would result in an extra 1.945 years of schooling to calculate the long-term benefits. Yeh evaluated the wider benefits of the program including increased annual income, the value of crime reduction, welfare and tax savings. The benefits were calculated from previous research or government documents. Discount rates were applied, and income calculations were based on a 44-year working life. There were five sensitivity analyses conducted on the effectiveness, costs, the robustness of the results and the income assumptions. The results are split into two strands of savings; first that the national social benefit-cost ratio is $28.47 (£32.27) for every dollar invested. For the

---

[8] Yeh (2009a), Bowden et al. (2015), Bowden and Belfield (2015a) all use the terminology BCA rather that CBA. As outlined in the introduction this is interpreted as a difference in language rather than technique.

federal treasury, the benefit-cost ratio is $93 (£105) for every dollar invested. There is a table in the analysis that breaks down the benefit-cost ratios for each state.

Belfield et al. (2015) conducted CBA on four interventions that were focused on social and emotional learning (SEL). They discuss the difficulties in collecting outcome data with this type of learning and appraise the use of CBA in the area of SEL, highlighting where adaptations from the conventional methods of CBA need to be made. The four interventions were 4Rs, Life skills training, Second Step and Responsive classroom. The data collected were from previous research. Outcome measures were different across all the interventions. A well-informed outcomes map demonstrates how outcomes can be aligned and provides a useful discussion on the way they applied shadow prices for each outcome. They also demonstrate how the benefit-cost analysis framework is applied to SEL. Extensive sensitivity analysis was performed on each intervention. The results are presented per 100 students with a lower and upper Net Present Value (NPV)[9], with a higher estimate and lower estimate for each intervention. For 4Rs, the NPV per students was between £226,762 and £13,497. For Life Skills Training, the NPV per 100 students was between £563,304 and £900. Second Step showed an NPV of between £2,386,799 and £75,587. The results for Responsive classroom were an NPV of between £1,885,184 and £464,692. The study aimed to demonstrate the use of BCA for SEL; however, the authors discourage the use of comparative NPV as the population outcome measures were heterogenous and ".it is not obvious that the most important impacts have been shadow priced." (Belfield et al., 2015, p.536).

Bowden et al. (2015) conducted a BCA on the intervention of City Connects. This intervention provides support for pupils based on the needs identified by teachers and school coordinators. This is a comprehensive intervention that is not solely targeted at attainment; the pupil's family and health are also part of the assessment criteria. Effectiveness data has been collated from previous research. Retrospective costing of the program was calculated using the ingredients method. Each ingredient was then inflated to 2013-dollar prices. Costs were at state level rather than national prices. The authors did not estimate wider benefits other than the reduction of dropouts. The total cost per pupil was £4112 and the benefit (mainly lifetime earnings related to not dropping out of school) was £12,461. Thus, the benefit-cost ratio was 3 dollars (£2.39) returned for each dollar invested. Furthermore, the authors suggest further investigation is needed on site variability to obtain more precise cost

---

[9] Net Present Value (NPV) is one of the ways in education that CBA is presented. In the simplest form the benefits are subtracted from the costs to give an NPV of the intervention the "profit "of the program. Benefit-Cost ratios, Internal Rate of Return (IIR) and Break- Even Analysis can also be presented Levin et al., (2018)

details and the other benefits that were not part of the analysis, as this could demonstrate more precise benefit-cost ratios.

Bowden and Belfield (2015) conducted a BCA as part of the evaluation of Talent Search. This intervention is targeted at supporting disadvantaged pupils to enrol in college or postsecondary education. The main outcome indicator is high school completion and post-secondary education enrolment. The authors used the lifetime model of earnings from Karoly (2015) and earning profiles from the population survey to calculate the long-term benefits of the intervention. Costs were collated retrospectively using information from senior leaders across the different sites, together with costs from publicly available documents. The authors used the ingredients method. The total cost was £23,036,124, and the net benefit £118,869,739. Thus, the benefit-cost ratio was $5.2 (£4.78) per $1 invested and for postsecondary the benefit-cost ratio is $54 (£47,81) per $1 invested. The authors conclude that there are difficulties in collecting cost data and that there is large site variability, and this warrants further investigation.

Table 3. 2 Main characteristics of included studies.

Studies	Yeh (2007)	Yeh (2009a)	Belfield et al. (2015)	Quinn et al. (1984)	Reynolds et al. (2002)	Bowden and Belfield (2015)
Country of origin	America	America	America	America	America	America
Type of intervention	Rapid Assessment verses, Voucher Programs, Charter schools, Accountability, and increased spending.	Rapid Assessment.	Social and emotional learning interventions, Responsive Classroom, Second step, Life skills training, 4Rs.	GEMS Math and Text Math.	Chicago Child-and Parent Centres that include a school age program.	The effectiveness of the Talent Search TRIO program.
Aims	To calculate which program is most cost-effective.	To calculate the Benefit-Cost ratio.	Demonstrate the use of Benefit- Cost analysis and the methodological implication of its use with Social and Emotional Learning.	To calculate which program is most cost effective for pupil's background.	To assess the wider benefits of the program at age 21.	To identify the cost effectiveness of the program across the different sites and Benefit-cost is also performed.
Type of population under study	All age groups in primary and secondary education	Age 6 to age 17	Each intervention is targeted at different grade levels. Targeted at different students and delivered in different ways	Grade 5 (10-11 years)	3 to 9.	Middle school or high school (11-18)
Type of evaluation	Cost- Effectiveness Analysis	Benefit Cost Analysis	Benefit Cost Analysis	Cost Effectiveness Analysis	Cost- Benefit Analysis	Benefit-Cost and Cost-Effectiveness Analysis.
Outcome measure	Rapid assessment STAR reading and Math test.	STAR reading, Math test and AFQT.	4Rs and Responsive classroom had educational outcomes on standardised tests. The other programs did not. Other outcome measures used either student, teacher or parental self-reported measures.	Two outcome measures were employed both are standardised tests, Iowa Test of Basic Skills (ITBS) and district level test (DMT)	High School dropouts averted	High school completion and pre-enrolment in collage. For the outcomes measure for Benefit- Cost Analysis the lift-time model of high school failure from Karoly (2015). Earning profiles were from Current Population Survey

Cost per pupil	$21.27 per year per student or $22.27 when including the opportunity cost to start up. This is for rapid assessment.	$22.27 or $28.31 when adjusted for the opportunity costs	4Rs per pupil $680 LST per pupil $160 SS per pupil $440 RC per pupil $2,160	GEMS $376 * Text Maths £448* *cost per class	Preschool program $4,400 per pupil per year* School age program$1,580 cost per pupil per year * *Programs were more than a year long, average per year is presented.	$680 pooled estimate across all sites.
2019 Price in £	£15.38 - £16.77	£16.77 - £27.03	4Rs per pupil £511.57 LST per pupil £120.63 SS per pupil £ 331.74 RC per pupil £1626.29	GEMS- £845.25* Text Math £1,097.88* *Cost per class	Pre-school program £5,056.45 per pupil per year* School age program£ 1816.06 cost per pupil per year * *Programs were more than a year long, average per year is presented.	£511.57 pooled estimate across all sites.
Perspective	Societal	Societal	Does not state	Does not state	Taxpayer	Societal
Main findings	That rapid assessment is more effective and less costly than the other interventions.	That Rapid Assessment would have a positive social benefit in terms of the investment into the program.	That the method of Benefit-Cost can be applied to Social and Emotional Learning.	That GEMS Math was more cost-effective than Text Math for higher SES pupils the reverse was observed for the lower SES pupils. This was across both outcome measures.	That across all the age groups there was a positive effect on the economic return to society.	There was a positive Benefit Cost ratio for both high school completion and post-secondary enrolment.

| Conclusions | That given the issues of criticism and implementation concerns this research demonstrates that Rapid assessment is more effective and less costly than other interventions that are targeted at raising pupil attainment. | That Rapid Assessment has twice the Benefit-Cost ratio than other research that has been carried out on raising student achievement. That the widening of the program to grade 1 through 8 would have a bigger impact and benefit every student. | That the method of CBA can be used but with the caution and recommendations for future research including the further development of the method. Benefits and the use of shadow pricing are not consistent, that the use of a benefit map is important. That the benefits presented are underestimated and no monetary value was found for some benefits. There should be caution when comparing Benefit cost ratios because of the missing benefits with no value. | The authors suggest including a measure of pupil ability in further research. Prior standardised measures of pupil's maths scores would have been beneficial. Further research is needed to address the variation in implementation costs of GEMS Maths. | This type of research and intervention are examples of high-quality interventions with evidence of long-term benefits for low-income children. Governments can use this type of research to make informed decisions on investing in early education. Even though money is invested into public school, Chicago Parent Centres are cost effective and have a positive cost benefit ratio. | That more information is needed on what students would receive in the absents of Talent Search, to increase the robust conclusions of the incremental costs and effects. Greater detailed information is needed on the variation across sites, and caution should be used when interpreting evaluations of this type due to the heterogeneity across sites. |

Table 3.2 Main characteristics of included studies continued.

Studies	Hollands et al. (2014)	Levin, et al. (1987)	Hollands, et al (2016)	Yeh (2009b)	Borman et al. (2002)	Bowden et al. (2015)
Country of origin	America	America	America	America	America	America
Type of intervention	Interventions aimed at high school completion. Talent Search, Jobs Corps, JOB START, NGYG and New chance.	The cost effectiveness of Computer-Assisted Instruction, compared with Peer mentoring, increased instructional time and Class size reduction	Cost effectiveness of early reading programs, Corrective Reading, and Wilson Reading System.	Cost effectiveness of Class size reduction and Rapid Assessment.	Cost effectiveness of Success for All, compared to Tennessee STAR, Perry Preschool and Abecedarian project.	City Connects an elementary and middle school intervention targeted to pupils who need further assistance.
Aims	To show how Effectiveness –Cost Analysis research can be used to inform policy makers in education.	To calculate if CAI is cost - effective compared to other alternatives.	To demonstrate the use of Cost - Effectiveness Analysis research in early reading interventions.	To evaluate the most cost-effective use of resources between two courses of actions.	To promote stronger links between school and home. To support students in aspects of behaviour, health and social issues that affect educational achievement.	To calculate the Benefit cost of City Connects using three different models.
Type of population under study	Each intervention was targeted at different age groups. The range covered Middle school (11-12) up to 24. Depending on the intervention.	Elementary pupils (5-10) Third grade pupils (8-9) Fifth grade (10-11)	Third grade (7-8) Second grade (8-9)	Kindergarten to grade 3 (5-9) Ages 6 -17	Pre-kindergarten – Grade 3 (4-9) Birth to Grade 3	Kindergarten – grade 5 (5-11)

Type of evaluation	Cost- Effective Analysis	Cost-Effective Analysis	Cost- Effective Analysis	Cost-Effective Analysis	Cost-Effective Analysis.	Benefit Cost Analysis.
Outcome measure	High School diplomas (HSD) or General Education Development (GED) and employability are the main outcome measures.	Does not detail the measure used only reports the effect size.	WRMT-R Word Identification subtest (a measure of word reading accuracy), WRMT-R Word Attack subtest (a measure of decoding accuracy), TOWRE Sight Word Efficiency subtest (a measure of word reading efficiency), and TOWRE Phonemic Decoding subtest (a measure of decoding efficiency).	Outcome measures from standardised testing (no detail of which tests were used) based on the results of Greenwald et al.'s (1996). Outcome measures for Rapid Assessment was *STAR reading* tests. Standard deviations were used for the effect size	Outcomes from Comprehensive Tests of Basic Skills (CTBS) to compare the intervention and the control group.	Outcome measures were high school dropout rates using maths and ELA in grades 6-8. They used propensity score model to calculate the high school graduates.
Cost per pupil	Talent search $ 3,290 NGYC $ 14,100 Jobs Corps $ 22,290 JOBSTART $10,460 New Chance $17,820	Cost per pupil was $119 for CAI	$6,696 for Wilson Reading System. $10,108 for Corrective Reading.	$21.72 per year per student or $22.27 when including the opportunity cost to start up.	$3,054 for Success for All	$4,570
2019 Price in £	Talent search£ 2,230 NGYC £9,555.71 Jobs Corps £15,097.42 JOBSTART £7089.60 New Chance £12,078.42	Cost per pupil £175.71	£5,107.47for Wilson Reading System. £7,709.52 for Corrective Reading.	£15.38per year per student or £16.77 when including the opportunity cost to start up	£3,509.19 for Success for All	£3,440.24
Perspective	Societal perspective	Does not state	Does not state	Does not state	Does not state	Does not state

Main findings	Talent Search is cost effective based on the other 4 interventions that are aimed at reducing high school dropouts.	Peer tutoring is the most cost effective for raising achievement for maths. Peer tutoring and CAI was almost equal for reading achievement. Peer was the most cost effective over all other interventions.	The Wilson reading intervention was more cost effective than Corrective Reading intervention but not on all levels of SES.	That Rabid Assessment is the most cost effective of the 5 other interventions.	That Success for All is more cost effective than the other intervention under investigation, even with the reduced cost of the other interventions.	That City Connect yields a return of around $3 for the investment.
Conclusions	There are difficulties in comparing the interventions due to the difference in delivery and target populations. Cost data was not comprehensive. The authors discuss this method of evaluation is important for educational interventions to support decision makers, but caution is needed when interpreting the results.	Caution is needed when interpreting results as the economic evaluation is only focused on one version and implementation of the intervention.	That cost effectiveness is a method that can be used in educational interventions that support decision makers. Call to the funders of research to financially support this type of evaluation and thus effectiveness data can be compared with other research to give decision makers accurate information for allocating resources.	The author discusses the hawthorn effect of Rapid Assessment, and that caution needs to be made when interpreting the results, the author also discussed the issues with the effect sizes of the other interventions. Suggestion that Rapid Assessment should be investigated further by funders as the results demonstrate robust cost-effective results.	The author discussed that this research supports Ramsey and Ramsey (1998) development theory. They suggest that rather than choosing one intervention over another, each intervention is needed as a risk presents itself, they call for expanding each intervention as it caters for a different need at a different time.	The results are based on conservative assumption and that the effects could be larger and that the cost could be significantly smaller.

3.4.2 Synthesis

Data was extracted from the included studies using Cochrane Review Guidance (Higgins & Green, 2011). A meta-analysis was not possible because of differences in population, outcome measures, and study designs.

The results from this review present the findings from full economic evaluations that have been carried out in education. CEA was the most common economic tool used for evaluating educational interventions, spanning eight of the included studies, with CBA/ BCA utilised in four of the studies. Given the disparities in the applications of the methods and the difference in the quality and design of the studies, it is difficult to draw any conclusions on which interventions are the most efficient at raising pupil's attainment.

2.4.3 Narrative synthesis

All studies had pupils of school age within their intervention groups. Reynolds et al. (2002) and the Chicago Child-Parent Centres evaluation is predominantly focused on pre-school intervention but had a school-age element, so it was deemed appropriate for this systematic review. All the interventions, except for Levin et al. (1987), targeted disadvantaged students who were either economically disadvantaged or risked academic failure. The age ranged from 5 to18 years of age for interventions with some of the comparator interventions supporting pupils up to 24 years of age.

The interventions included in the studies focused on a range of academic and well-being outcomes, e.g., Yeh, 2007,2009a, 2009b reported mathematics and reading outcomes, Quinn et al (1984) compared two mathematics programs, Hollands et al. (2014) explored reading interventions, Levin et al. (1989) focused on CAI, and Belfield et al. (2015) identified SEL interventions. All these interventions can be grouped as smaller-scale school-level interventions. The heterogeneity between the interventions prevents a useful meta-analysis and/or meaningful conclusions to be drawn.

There were some similarities in the choice of comparator. Levin et al. (1987) Yeh, (2009a), and Borman et al (2002) used class size reduction as one of the comparators. Levin et al (1987), Yeh, (2007), Hollands et al. (2014) Borman and Hewes (2002) used multiple comparator interventions to determine the cost effectiveness of the intervention under analysis. Comparators were large policy intervention and based on previous research. Quinn and Mondfrans (1984) and Hollands et al. (2014) used similar scale interventions for their analysis.

Outcome measures ranged from standardised tests to intervention-specific measures. All studies reported effect sizes to calculate the cost-effectiveness or future earnings in the cases of CBA Given that most of the studies used effect sizes from previous research in the form of standard deviations, there is not sufficient information presented to identify if the outcome measures are comparable.

None of the included studies calculated an Incremental Cost-Effective Ratio (ICER); instead, effectiveness-cost ratios were used in a league table of included comparators. Although the information was there to calculate an ICER should consumers want to take the analysis a step further. The disaggregated presentation might not be helpful to policy makers, within health economics ICERs are used in this way to rank interventions.

3.5 Discussion

The focus of this discussion will highlight the technical disparities in the application of economic evaluations within educational research. This will help to support decision makers to interpret the available data and provide some additional thought for future evaluations and research. As shown in Table 3 1, the quality of the studies included within this review ranged from 3 to 10 suggesting a range of poor quality and good quality economic evaluations being carried out in education. Common themes were identified and are discussed below in terms of the methodological and measurement issues, the choice of comparator, and cost-effectiveness calculations. Whilst the existing data provides some insight into the cost-effectiveness of these interventions in education, education research more generally would benefit from higher-quality evaluations to assist in decision making, as with the case in health discipline (Henrikson & Skelly, 2013). This will help to revert to the central issue of supporting decision makers in identifying research that supports academic achievement and value for money.

3.5.1 Cost data

Most of the papers used Levin's (1988) ingredients method to collect cost data and provided uniformity in collecting of costs. Yeh (2007) gathered costs information based on a variety of assumptions (from the number of students to teacher turnover) and failed to provide a clear description of the methodology employed. Yeh (2007) also uses opportunity cost as a sensitivity analysis in terms of the time that the teacher takes to monitor the student. As a standard part of the intervention, this cost should be included in the main analysis. If monitoring of pupils is required, then varying the staff member (e.g., teaching assistant or

fully qualified teacher) could be incorporated into the sensitivity analysis. With these issues it is difficult to draw conclusions given the inaccuracy of the costing information.

Bowden and Belfield, (2015) did not incorporate the cost of training teachers or the post-secondary cost of running the project. Given that these are costs associated with the running of the intervention, Levin et al. (2017) suggest that all costs should be incorporated to avoid underestimating the true cost to run an intervention. Given that training teachers and the personnel time to run the intervention would incur costs, this could affect the overall cost effectiveness analysis and the benefit–cost ratios presented for the intervention. By not calculating all the costs, decision makers are likely to underestimate the true cost of an intervention or program.

Retrospective costing was a common theme. However, using state or government data or national prices can lead to uncertainty in local pricing, distorting the true costs of interventions. An example of this is the Bordman and Hewes (2002) study where a reduction in the cost of the comparator by the proportion that the intervention could have affected mathematics and reading, could provide misleading information to decision makers. Given the limited scope of economic analyses that have been carried out in education, retrospective costing is a promising way forward for decision makers. However, costs are more precise when collected contemporaneously with the intervention or program (Hollands et al., 2016; Levin, 2002).

All the papers included in this review cite Levin's work on collecting cost data using the ingredients method and all cite his work on cost effectiveness methods in education (see Levin et al., 2017). They conclude that a consistent framework is required to ensure comparable cost data are collected. Greater sensitivity analysis on pricing and assumptions of cost would also enhance the credibility of the results. And as Bowden et al. (2015) also suggest that researchers should avoid making comparisons of economic evaluations that do not use preferred accounting methods such as the ingredient's method. The authors also note the importance of collecting costs contemporaneously with the intervention or program to improve the quality and robustness of any subsequent economic evaluations.

3.5.2 Measures

None of the studies reviewed here utilise a consistent academic and/or well-being outcome. The studies used a variety of education measures, e.g., high school completion rates, mathematics and reading outcomes, and standardised tests were common. Some of the papers only report effect sizes in relation to the comparators. In order to compare

interventions, it is important that researchers provide information on the measures used for both intervention and control groups. Clune (2002) review also identified issues of comparability, pilot studies for example being compared to larger scale policy intervention are not helpful to policy makers due to the generalisability of finding from small scale research.

Yeh (2007, 2009a, 2009b) used a program specific measure and the resulting effect size could be replicated on the Army Forces Qualification Test. With no mapping or justification to compare the two tests or that the results can be replicated, limited conclusions can be drawn. Kraft (2020) details that program specific measures have 2 to 4 times higher effect sizes than standardised tests. Also, Kraft (2020) suggests tests administered immediately post intervention can yield greater effect sizes than tests administered following a longer period of time. Regarding the issue of whether standardised tests or program specific test are used the researcher needs to be sure that it is appropriate to compare them; the effect size is not enough information especially given that the premise of CEA is that outcomes should be the same or similar to allow meaningful comparisons.

Cost-effectiveness Analysis: Ratios, ICERS and Cost-effectiveness Planes

A consistent theme was the difference in the calculation of the ratios. Most of the papers in this review present effectiveness-cost ratios, the only papers that used the cost-effectiveness ratio were Hollands et al. (2016) and Quinn and Mondfrans (1984).

Within health economics, cost-effectiveness ratios are calculated using the method described by Morris et al. (2012) as follows:

Cost-effective ratio: The cost is divided by the effects.

$$\frac{costs}{effects} = CE\ ratio$$

Effectiveness cost ratio: The effects are divided by the costs (Levin et al., [2017] for an example of the different results see the following footnotes).

$$\frac{effects}{costs} = EC\ ratio$$

The papers that detail effectiveness-cost ratios extrapolate results against fixed dollar benchmark costs. This allows decision makers to see what effect size can be purchased for example $100,000. Table 3.3 illustrates the disparities of how the results are reported and the extrapolation of the results to different dollar amounts. Within the effectiveness-cost ratio

papers, the results were ranked along with alternatives. For clarity, the effectiveness cost ratios have been converted into cost-effective ratios in the final column.

Given the disparity in the way that the ratios are calculated and presented it would not be unreasonable to rename the method as cost consequence. Within the cost consequence method, costs and outcomes for each intervention are presented in a disaggregated way that allows the decision maker to make their own judgments (Mauskopf et al., 1998). This is an issue of clarity; Levin et al. (2017) suggest that although some education research does present effectiveness-cost ratios, given that the recommended standard in health evaluations is to present cost-effectiveness ratios, and to avoid confusion education should do the same. The extrapolated results could be misleading if the whole cost of the program is not clear.

Table 3. 3 Conversion of effectiveness-cost ratios to cost-effectiveness ratios

Paper	Cost per pupil	Effect size	Number of dollars for effect size	Ratio	Cost-effective ratio
Yeh, 2007	22.77	0.319		Effectiveness to cost ratio 0.01432420	$70 per effect size gain
Quinn et al., 1984	$288 for GEMS Math $194 Text Math	GEMS Math Text Math		Cost effective ratio $11.4 per raw score (ITBS tests) $23.72 per raw score (DMT) $13.45 per raw score (ITBS tests) £81.07 per raw score (DMT)	
Bowden et al., 2015	$43,440 per high school completion	High school completion	$100,000	Effectiveness to cost ratio 2.3 extra graduate	$43,440 per effect size gain
Hollands et al., 2014	$30,520 per high school graduate	High school completion	$100,000	Effectiveness to cost ratio 3.3 extra graduate	$30,520 per effect size gain
Levin et al., 1987	$119	0.177	$100	0.15 effect size [10]	$680 per effect size gain
Hollands et al., 2016	$10,108 Corrective reading $6,332 Wilson Reading System	0.22 Corrective reading 0.33 Wilson Reading System		Cost effective ratio $45,945 per effect size gain.	
Yeh, 2009a	$ 28,31	0.319		Effectiveness to cost ratio 0.01126810	$88.75 per effect size gain.
Borman et al., 2002	$3054	Reading 0.29 Maths 0.11	$1000	Effectiveness cost ratio^2 0.09[11]	$ 10,531 per effect size gain

---

[10] Levin et al (1987) calculates an effectiveness cost ratio for the research on CAI
CER would be 119/0.175 = $680
ECR would be 0.175/ 119 = 0.00147 * 100 = 0.147 (0.15)

[11] Borman et al 2002 reports a cost-effective ratio per $1000 however he uses the effectiveness cost ratio e.g.
CE ratio would be 3054 / 0.29 = $ 10,531 per effect size gain
EC ratio would be 0.29/ 3054 = 0.00009496 then x 1000 = O.09 effect size for $1000

Cost-effectiveness Planes

Bowden et al. (2015) uses a cost-effectiveness plane to display the simulated results across the sites, the sloping line displayed is the cost-benefit test (forming the acceptability line); the cost-benefit test is simply that the intervention does not return a negative ratio. The use of the cost-benefit test as a threshold or acceptability line is not explained in the paper. Economists and decision makers usually work within a fixed budget or a celling threshold where the acceptability line is drawn. Therefore, the use in this research is of limited help to aid decision-making for those without a background in economics.

Incremental Cost Effectiveness Ratios (ICERs).

None of the papers attempt to calculate Incremental Cost Effectiveness Ratios (ICERs) as part of their cost-effectiveness analysis. This is a calculation that shows the best value alternative between interventions. An ICER[12] is the difference in the costs of each intervention and then divided by the difference in the effects. Using an ICER is relatively common in health research, and this allows the ranking of interventions when outcomes are the same or similar to aid decision makers, ICERs can also be used to determine whether interventions meet the cost-effectiveness thresholds for NICE guidelines or Health Technology Assessment (HTA) (Jayasundara et al., 2017.).

Yeh (2009b) uses a Relative Effectiveness Cost Ratio (REC)[13] which has similarities to an ICER value in that it compares costs and effects of competing programmes. Whilst an ICER displays the incremental cost to gain one additional unit of effect when investing in programme A over B, the REC displays the relative size of effect that could be bought for £1 when investing in programme A over B. Yeh (2009a) is the only study found by this review to use a REC; as is not mentioned in any of the health or education literature it's application and relevance could be considered limited.

Yeh (2009a) argues that this ratio can be used to compare interventions. As mentioned previously, although the methods are similar, within the health discipline RECs are not used and are not commonly cited within the literature. Although Hollands et al. (2016) discuss the

---

[12] Incremental cost effectiveness ratio:

$$\frac{cost\ of\ A - cost\ of\ B}{effect\ of\ A - effect\ of\ B} = ICER$$

[13] Relative Cost-effectiveness Ratio:

$$\frac{effect\ A/cost\ of\ A}{effect\ of\ B/cost\ B} = REC$$

use of ICERs, their study does not present sufficient detail to allow wider conclusions to be drawn. These studies provide useful information and examples of how ICERs are calculated. Given that some of these studies are comparing multiple interventions, the addition of ICERs would further transform the accessibility to decision makers (see Levin [2017, p.180] for an example).

3.5.3 Choice of Comparator

CEA within health research has always faced the difficulty of identifying comparators that are either similar or have the same outcome. While the health discipline has sought to mitigate the 'apples and oranges' debate, education research has yet to fully embrace the complexity surrounding the debate. Health economists have developed generic questionnaires so that outcomes can be compared even if the units or measurements are different; this amounts to Cost Utility Analysis which is a type of CEA (Morris et al., 2012).

Only one of the papers (Hollands et al., 2016) discuss why or how they chose the comparator for analysis. Many of the included studies were retrospective which might have restricted choice. There are particular issues with selection bias, for example Levin et al. (1987) and Yeh (2007) used previous research to build the cost effectiveness data. However, they do not clearly state what criteria they used to select the studies for comparison other than interventions that were targeted at similar outcomes, populations, scale, implementation fidelity and outcome measures are not discussed. Borman and Hewes (2002) took a pragmatic approach, by reducing the cost of the comparator to try and demonstrate what effect could have been linked to mathematics and reading. The argument against this choice of comparator would be that from the outset the interventions target different range of outcomes. This approach created uncertainty in the cost and effects for the comparators and the overall choice of an appropriate comparator.

Program scale is an issue when considering the comparability of interventions. Costs are often not directly comparable, for example with policy wide intervention compared to small scale interventions as in the case of Yeh (2007, 2009a, 2009b). Small scale interventions can seem cost-effective, relative to school/county wide interventions as Kraft (2020) suggests "Policy makers have to consider not only what works but how it works relative to the costs and the financial investment required" (p. 246). Levin (1987) suggests caution should be made when interpreting the results as different applications could yield different results when generalising to different education systems, as is the case with health research (Salkeld et al,.1995). Context, both domestically and nationally, should be in the

forefront of the decision makers mind, all too often context of research is not questioned and limits the transferability to other contexts (Little, 1996). Also, researchers need to consider that scaling of research primarily depends on the program specifics and, secondly, the context of where the program may be scaled to. Issues of the Hawthorn and Rosenthal[14] effects are research specific issues, but when looking at scalability, area demographics, and heterogeneity in students' needs to be considered, "understanding of the contextual variables that matter" (Sternberg et al., 2011, p. 20). For an in-depth discussion on the issues or scalability and generalisability in educational research, see Sternberg et al. (2011).

Holland et al. (2014) comments that interventions vary in many ways and suggests that it limits the comparability of them. They discuss decision makers' need to be cautious in interpreting the results and consider context and target population within CEA. Education and health research face the same issue. The choice of a comparator for cost-effectiveness is an important issue, especially if results are used to make financial decisions on a large scale. With CBA, comparisons of different interventions need to be transparent to what is comparable and what is not. Belfield et al. (2015) suggest that the use of a benefits map can assist decision makers in understanding where interventions are comparable and where they are not.

When choosing a comparator for the interventions, researchers must be mindful of what the outcome indicators are, the scale, context, population. While decision makers also need to be aware of the context of the intervention, for example a tightly controlled intervention in a school compared to a large-scale behaviour change study across a group of schools.

3.6 Strengths/Limitations

This systematic review took a health perspective using a health quality appraisal tool. Although this was an appropriate tool for this study, other quality appraisal tools are available (see Husereau et al., 2013). Given the epistemological stance adopted, the appraisal tool used could have led to an overly negative appraisal of included studies and thus a limitation of this review. While critical appraisal tools for educational economic evaluation are not currently available and given that education researchers have looked at the discipline of health research

---

[14] The Hawthorne effect is when participants know they are being observed thus modify behaviour and this in turn can limit the research (McCarney et al., 2015). The Rosenthal effect proposes that high expectation will yield better results and that low expectation will yield lower results also known as the experimenter effect (Colman, 2015).

(Levin, et al., 2017) it is considered appropriate that a health economics critical appraisal tool was used.

A limitation of this review is that other languages were not considered. All the papers were from the United States and are in English. Papers are also from a narrow field of researchers; Levin, Holland, Belfield, and Bowden are all researchers working from the Centre for Benefit Cost Studies in Education (CBCSE). They make up half of the included studies and are ranked the highest quality economic evaluations. There is a lack of diversity in the application of the methods.

We only searched academic databases for this review. This could present a possible bias in the results as some policy evaluations are published in the grey literature. The focus of the review incorporated research within schools; however, there are examples of economic evaluations in higher education, pre-school and residential or specialised education settings.

Although this review identified economic methods used in education research, a large amount of this research fails to provide a full economic analysis to support the conclusions that are drawn. It is important that researchers and decision makers are aware that full economic evaluations are currently rare in the research literature, and that caution should be applied when consuming research claiming to be cost effective, for example, see Salkeld et al. (1995) for a similar discussion of health research.

3.6.1 Implications for policy and practice

While decision makers grapple with the 'what works' and 'how much' this review could bridge the gap between research and policy decisions

Clarity in reporting of outcome measures would enhance not only the comparability of interventions but this would enhance decision makers understanding of the interventions under investigation. Outcome indicators as well as the effect size, enable greater clarity.as to whether the indicator is reliable, comparable and suitable for the intervention. The type of outcome indicators, age of participants and whether the effects were casual or correlational are issues that have been highlighted for educational research (Kraft, 2020).

Both education and healthcare face broadly similar challenges of scarce resources and complex populations. Levin (2002) describes research in education as often idiosyncratic, thus posing considerable challenges for policy makers to consume this information and make informed decisions. It would be advantageous if the disciplines worked together to share expertise.

Consumers of research are advised to question research that claims to be cost effective, researchers should also be cautious labelling interventions cost effective without a full economic evaluation. Simplistic calculations, or even assumptions, do not provide the in-depth information that is needed to either upscale interventions or make informed policy decisions.

Cost data needs to be collected routinely when evaluation takes place; retrospective estimation of costs can lead to uncertainties and assumptions, which without robust sensitivity analysis can lead to misinterpretation. Transparent cost data would provide decision makers with up to date and more precise costs for interventions. Kraft (2020) suggests that increasingly research in education does not include this type of cost analysis and the use of rudimentary cost estimates is commonplace. Costs data collection is incorporated very early in an intervention's design or in trials of treatments in health research. 'Piggy-back' economic evaluation can assist decision makers early in the research process (see O'Sullivan et al., [2005] for further discussion in relation to health research).

The use of CEA ratios is advised, and this is in line with the recent literature Levin (2018), the presentation of effectiveness-cost ratios that are extrapolated to certain dollar amounts could be misleading to the total cost of interventions. The regular reporting of CEA ratios would allow comparisons to be made. In addition, calculating an ICER may be more beneficial to decision makers than extrapolating results for certain money values as seen in the cost effectiveness studies in this review, even if the comparator is normal provision. Ranking ICERs can assist decision makers to make the best use of budgets (Bilinski et al., 2017).

One implication for research would be a development of a common indicator of effectiveness, as seen in the health discipline in the terms of a QALY, this allows the comparison across different diseases or interventions. This would alleviate some of the issues when finding comparators in educational research, a QALY is by no means a perfect measure (Neumann et al., 2018) but does provide good information for decision makers in health and especially when used with thresholds. But only focusing on the activities that yield the biggest impact on one particular metric – attainment and GSCEs as an example- risks reducing education to a narrow and one-dimensional outcome (Levin, 2012). This is the criticism of using the QALY in health research, for example, some commentators in America suggest that a price cannot be put on life, like the QALY thresholds in UK. But also in education, the risk is not understanding the importance of non-cognitive skills can have on outcomes. But also in education, the risk is not understanding the importance that non-

cognitive skills can have on outcomes. Researchers developing an effectiveness measure for education would need to consider balancing the efficiency of a universal metric against the potential loss of sensitivity that may result.

Finally, it would be advantageous to the discipline of education economics if there were a critical appraisal tool that was specific to economic evaluations in education, the use of a health economics appraisal tool could have limited the findings given that Drummond et al., (2015) is specific to health. A specific tool for education would support the best practises within education research and would allow consumers of education research to make more informed decisions about the quality of the research.

3.7 Recommendations

1. Before commencing efficacy or effectiveness trials, researchers should consider costing the intervention even at the early stages.
2. In the case CEA evaluations, the choice of comparator is an important factor to ensure that meaningful conclusions can be drawn between studies. Outcome indicators and scale are to be considered when comparing interventions.
3. Cost-effective ratios should always be presented to ensure uniformity in the discipline of education economics.
4. Where possible, a common indicator of effectiveness would enhance the comparability of interventions.
5. While this review used a health economics critical appraisal tool, education research that uses economic analysis would benefit from developing a bespoke tool to strengthen the quality and standards needed for education economic evaluations.

3.8 Conclusion

This review presents several important issues related to the use of economic evaluations in education. Evident is the need for more, while this review and Clune's (2002) did initially identify many hundreds research papers that were claiming to be economic evaluations this was sadly not the case. More comprehensive economic evaluations should be undertaken to support decision makers more effectively on how to improve outcomes in

schools, including reducing the attainment gap between more and less advantaged learners. Furthermore, education research needs to improve how it identifies and collates information on program costs, researchers in education should take note of the efforts in health where costing are collected very early in the trials of research but also as part of most research that is conducted in health. A key feature of this review is the need to inform decision makers about how to make the most efficient use of the scarce resources to improve outcomes for learners. As Godard et al. (2020) notes, improving educational research could save money that is wasted on ineffective strategies, research focused on education economics can support this movement towards evidence informed practice and policy. But there needs to be considerations on what value is, and not reduce outcomes to a single variable, which could be misleading to decision makers and drive education evaluation to too narrow a focus (Levin, 2012).

Chapter 4

Interviews and Focus s groups (Studies 2 &3)

"It's not the person who knows … it's the person who doesn't know." Teacher school R

4.1 Introduction

As a reminder, the FAIP evaluation was split into three empirical studies to build the evaluative framework. This chapter concentrates on Study 2 (the semi structured interviews with teachers) and Study 3 (the focus groups with pupils) in the intervention schools. Each study contains a direct discussion, with a general discussion at the end of the chapter.

Ethics

Ethical approval was granted for both studies from Bangor University Psychology Ethics and Research Committee application number :2018-16324-A14505 for both studies. Consent documents for interviews and focus groups can be found in appendix C1.

4.2 Study 2: Teachers interviews

4.2.1 Research focus and design

Qualitative interviews with teachers were conducted to explore the experiences and perceptions of formative assessment disseminated through a professional development programme. The teachers were trained in strategies by an expert trainer commissioned by the School Effectiveness and Improvement service North Wales (GwE). Continued support was provided by other project teachers as part of a train-the-trainer professional development model. Teachers were trained in various formative assessment strategies over an academic year. They were free to disseminate whichever strategies they wanted into their class (See Chapter 2 for a full and detailed description of the intervention design).

Study 2 focuses on evaluating teachers' perceptions of using the strategies, offering some insight into the wider evaluative framework and perceived impact. As the teachers are not passive actors in research, it is important to gain their perception within the Theory of change. This input greatly supports the generation of understanding about the phenomena under investigation (Blamey & Mackenzie, 2007).

4.2.1 Research questions

Study 2 aims to answer the following research questions

1. What are the perception and experiences of Tier two teachers in relation to implementing formative assessment strategies and the perceived impacts in the classroom?

2. What are the barriers and facilitators of implementing formative assessment in the classroom?

4.2.3 Participants

Teachers were recruited to be part of the FAIP evaluation (See chapter 5) to participate in one-to-one interviews. Semi-structured interviews were conducted with seven out of the nine intervention schools at the end of the 2019 academic year. Table 4.1 details the characteristics of the schools involved in the research. Due to illness two teachers in the intervention group were not able to participate in the interviews. Limited timescale meant the interviews could not be rearranged.

Details of the 7 participating teachers is shown in Table 4.1 and Figure 4.1 details the elements that the teachers trialled within the classroom over the academic year. As the teachers could implement any elements covered in the formative assessment training, there were some strategies that were evident prior to this implementation. An example of this is Growth Mindset (see Claro et al., 2016). Some of the schools had already implemented this in their classroom(s) so were able to focus on certain elements within it. Interviews were used to collect time and costs for teachers to run the project in their school this information was used in the cost consequence analysis presented in Chapter 5.

4.2.4 Procedures

Interviews were conducted in schools, mainly in a classroom or a quiet location. All interviews were recorded and later transcribed. A semi-structured approach was adopted, with questions focusing on the new strategies implemented, and whether there were barriers to implementation, including the impact teachers observed in their classrooms (see Appendix C2 for topic guide). Time was left for respondents to discuss further matters in relation to the project, and for the researcher to explore points raised that were of interest.

Table 4. 1 Characteristics of participating schools

Schools	Number of statutory school age pupils	Main Language of instruction	Local authority	eFSM %	Gender
School L	82	Welsh	Anglesey	8.5	Male
School M	83	Welsh	Gwynedd	19.3	Male
School N	179	Welsh	Gwynedd	34.6	Female
School P	326	English	Wrexham	23	Female
School Q	57	Welsh	Gwynedd	8.8	Female
School R	355	English	Flintshire	8.2	Female
School T	287	Welsh	Anglesey	29.3	Male
Mean	196			18.8	

Figure 4.1 1 New elements that the teachers implemented in the classroom
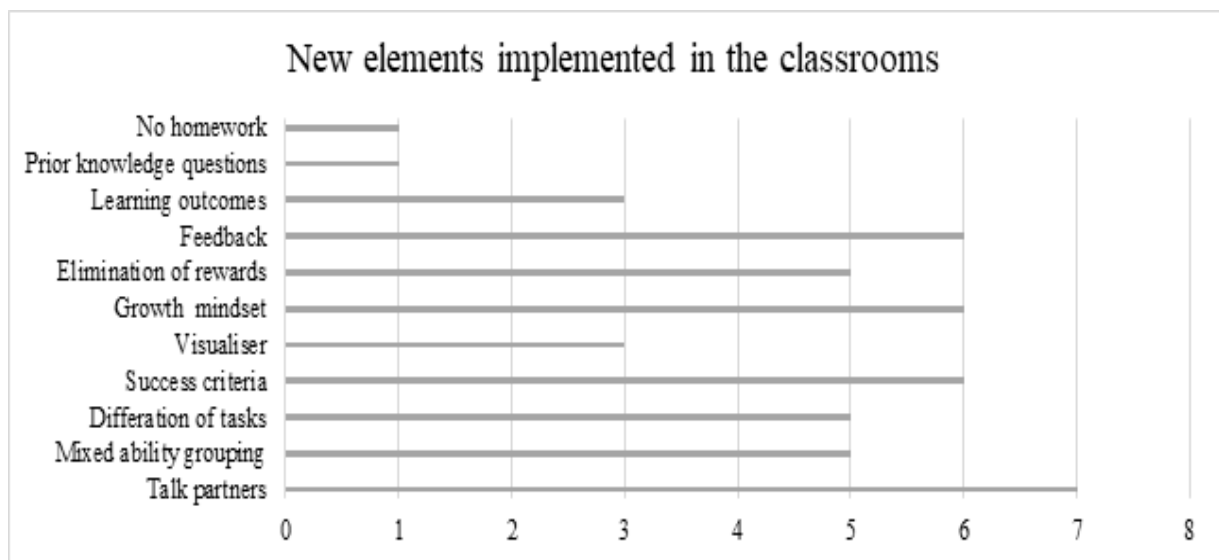
4.2.5 Analysis

The researcher used the six steps to thematic analysis to identify key themes across the transcripts (see Chapter 2 for a full discussion). This approach supported our ability to address the research question (i.e., explore facilitators and/or barriers) whilst leaving room for other concepts/phenomena to be discovered (Braun & Clarke, 2012).

4.3 Results

Two main themes were identified: positive factors associated with the programme (including the benefits to learners); and negative factors associated with the programme. Themes focused on facilitators and barriers will be discussed separately. The following 10 subthemes were identified, see Table 4.2.

Table 4. 2 Themes and sub themes

Theme	Positive factors	Barriers	Facilitators
Subtheme	Encouraging more independent learners	Resistance	Behaviour change
	Focused pupils on their tasks	Context	SLT Support
	Understanding of pupil progress		Stakeholders' involvement
	Self-efficacy		Adaptation
	Behaviour		
	Reduced workloads		
	Improved standards		
	Prepared for the new curriculum		

4.3.1 Theme 1: Positive factors associated with the programme

Encouraging more independent learners

All teachers detailed that the strategies they employed allowed the pupils to become more independent, they were able to not only get on task quicker, more ready to learn, but were able to self-assess and reflect on their work.

> *"Yes. And I think because of that, because they sort of plan and they guide themselves how they want to go, it has definitely had an improvement on their enjoyment and their willingness to work."* Teacher school R

"But, because of this, we're find that we've got less low achievers. They are much more independent. So, they know what to do." Teacher school L.

" So, I think by having those chillies[15], we are making sure that they are independent, and you know, that they're able to have the time to reflect back on their work." Teacher school T.

Focused pupils on their tasks

Teachers explained that the pupils were focused on what they were learning and were able to assess their own work as well as the work of the other pupils. This included being able to see the mistakes or areas of improvements. The use of the formative assessment strategies, particularly the success criteria and the bullseye, proved to be effective tool in helping the pupils to understand what was expected from the task and if the task was challenging them. Allowing them to think about the tasks they were given, and support where they are in their own learning and focus tasks.

" …. And then I said, 'If you're here, are you learning anything else?' and then she said, 'No, I'm not learning anything…. so, then she got up and changed her work." Teacher school Q.

"Whereas now they say "actually Miss N [Class teacher], I'm in the comfort zone here. I need something a little bit more challenging." Or they might start on a particular task, get on with it alright and then say, "I thought I was on the green chilli, but I'm in the comfort zone now after trying a few and I would like something to put me back into the learning zone." Teacher school N.

While pupils could use the strategies to focus their learning, the teachers were able to use feedback that was specific to the pupil so that they could understand what needed to be done with particular pieces of work. A teacher from school N discussed how a 5-minute conference with a pupil one to one gave them a better understanding of what they needed to focus on within their work, rather than the previous feedback strategy that would be written in the pupil's book. A teacher from school R explained that she was able to walk round the class with an iPad and leave verbal feedback for the pupils, and this allowed instant feedback, but also pupils could stop and play back the feedback at their own pace. This in turn

---

[15] Chillies are indicators to the pupils the level of difficulty in a task.

supported the pupils to be focused on their work as they received specific feedback on the task they were completing within the real time

Understanding of pupil progress

All teachers agreed that the strategies supported them to know where the pupils were in their learning and a better understanding of who the pupils are as individuals. Teachers were able to identify which pupils needed support much quicker and adapt teaching in the real time to support any misconceptions that pupils had within the learning process.

> *"I had allowed them to choose their level of work, which has meant they have been able to push themselves and if they've got specific learning issues, it's identified them more."* Teacher school P

> *"I find out at the end its wrong, they find out the next time they were wrong, and they try get it right in the moment". "With misconceptions you get a much better chance to call out a misconception." "That's been revolutionary. Because I would struggle, I think, to feel that I got to know all the children".* Teacher school R

> *"We would group them at the start of the year and within two weeks you would know. I expect these children to get Level 4, Level 5 and Level 6. Now we ... it changes quite often…. And you kind of ... you know things can change a lot, but the actual assessments, the summative assessments aren't made until later on in the year, which gives us a greater chance to allow them to develop and it turns out I've seen a shift in the sort of less able to the middle group and more of the middle group into the higher group".* Teacher school L.

> *"...helps me feel I get a greater understanding of my children. And I don't go home at the end of the week thinking " I don't think I've said five words to that child".* Teacher school R

Self-efficacy

Self-efficacy is the notion that pupils have positive feelings towards themselves and feel confident to achieve their goals (Leahy & Wiliam, 2014). Teachers detailed that there was an improvement of pupil's self-efficacy within the classroom—they were able to challenge themselves to do tasks they were hesitant at before. Teachers also found that pupils felt more able to speak out and share with the class.

"And sometimes it can be a little bit of idleness of picking up a pen but sometimes it's their belief in themselves a lot of the time. And it is, it's them thinking, "actually, I can do it". "I think a lot of it is the confidence they have…" Teacher school P

"Yeah, I've got an example in one of the maths lessons where one of my lower ability children had finished and then I had like a challenge, and then I asked, 'Would you like to do the challenge?' and he said, 'Ooh, no. No,' he said, like that, like. And I said, 'Have a go,' I said. Just… …have a go and showed him what the challenge was, and he did it and he was so chuffed with himself." Teacher school Q.

Two teachers detailed they had received feedback from parents that the new strategies had impacted their children positively. These children were more confident and appeared happier coming to school.

Behaviour

Every teacher that was interviewed commented on how the dynamics with the classroom had positively changed. As well as supporting better bonding between pupils the strategies also created more inclusive classrooms, so pupils were not left out. This was mainly due to implementing talk partners, meaning pupils were sat with different peers each week. All teachers modelled good talk partners. Pupils developed friendships, there was less arguing in the classroom and in the playground.

"I found there was less fighting, or less arguing outside …" Teacher school M

"Socially it's more inclusive…" Teacher school Q.

"But, because of that, then it's developed sort of other relationships within the class…. Because they are used to working with other children. You will then see them kind of playing outside with them. So, it reduced the amount of bickering and arguing we have. We find there's not so many cliques of children now…. I could probably honestly say last year we didn't have any children that felt left out. They always felt they could go to someone and largely I feel that's because of this". Teacher school L.

Reduced workloads

All teachers identified that the feedback was able to reduce their marking workload. This was mainly attributed to the improved standards of work, which in turn allowed them to

mark quicker. Pupils were able to use the success criteria to make sure that they included all the relevant elements within their work.

> *"So, it's easier. I think the quality of work is easier to mark……. I do feel I've got extra time."* Teacher school M

> *"…. we're getting through more……. timetable space we were able to bring things forward,"* Teacher school R

The teachers described that they were able to get round the pupils quicker and were not tied to sitting with a particular group. All the teachers suggested that implementation of the different strategies did initially take some extra time, but they were able to gain this back over the academic year. One teacher saved 45 minutes a day while another detailed an hour a day was saved.

Improved standards

The teachers all discussed that there was an improvement of standards, this fell in to two categories. Firstly, improved standards of work that was presented to them either with the use of success criteria, talk partners or peer and self-assessing. Secondly, teachers were able to see pupils have moved out of their predicted level, and this was particularly evident with underachieving pupils. The pupils were more likely to challenge themselves to try harder work.

> *"I would say predominantly it's that the lower achievers it's had the bigger impact on". "… I've … you know, we've found now that the less able children will … because of this and because of their growth mindset that we've trained them with, they're much more willing to try more difficult tasks. And because obviously then you have higher academic achievement …"* Teacher school L

> *"We did have a few more level 5s this year than we thought we would."* Teacher school Q

Prepared for the new curriculum

As Wales is undergoing major curriculum reforms at the time, teachers were asked if they felt the strategies that they were employing were conducive with the new curriculum. Teachers interviewed were able to provide detail on three different levels: pupil, strategic and preparedness. Pupils were becoming independent, ambitious, reflecting on their learning,

teachers were focusing the pupils on their work, problem solving and wellbeing. Teachers also detailed formative assessment linked with, the 12 pedagogical approaches and the 4 purposes set out in the Curriculum for Wales (Welsh Government, 2019a). While one teacher did not agree and felt there was not enough information about the curriculum, another teacher's school had begun to implement the new curriculum, some felt that this project has prepared them for the roll out of the new curriculum.

> *"So ... and I have looked at the new curriculum a lot. You know, looked through the different ... so I think all of this; the way they have to work alongside each other, the growth mindset, the wellbeing side of it ..."* Teacher school P

> *"Yeah, of course. The entire Shirley Clarke assessment for learning is something ..., so the 12 principles of learning or whatever it's called, they obviously borrow things from assessment for learning, don't they? For you to be able to use various different strategies in your classroom, all of which Shirley Clarke has discussed and people have trialled. It's only going to ensure that pupils are getting the best opportunities in class. For children to ensure that ... just looking at what a good one looks like, that ensures that children are being ... that they want to achieve ..."* and *"Moving in the right direction for the new curriculum."* Teacher school N

All teachers stated that the strategies they have implemented into their classroom changed the culture and dynamics, and formative assessment provided a positive impact for both the teachers and the pupils.

4.3.3 Theme 2: Barriers and facilitators associated with the programme

This section describes the barriers and facilitators that teachers experienced when implementing formative assessment strategies from formative assessment programme.

Barriers to implementation and dissemination were evident, both internally (classroom/ pupils) and externally (context, and stakeholders). Two barriers were identified: resistance and school ethos/context, and four facilitating factors were identified: involvement of stakeholders, adaptation, and teacher reflection.

4.3.4 Barriers to the programme

Resistance

Resistance was a common theme and related to different stakeholders. One teacher explained that they had to 'go it alone' with other teachers not supporting or willing to try anything new. The main resistance for teachers was eliminating rewards:

"And we did it for a few months, but it was difficult because we were the only class in the school that was doing it. Everybody else was still doing the ClassDojo points and I found it difficult…." Teacher school N.

Resistance from different departments was also identified, reward systems in the foundation phase were intrinsically linked to other educational theories and programmes.

"Because I think historically, they rely on rewards for … to motivate children really." Teacher school L

"*That's it. Webster-Stratton…"* Teacher school M.

While this different way of teaching was a source of conflict for the teachers, attitudes and mindsets also proved to be problematic both within the classroom and other departments,

"People who were Teaching Assistants last year … that was what was hard to say, "you need to step back". Teacher school P.

Teachers experienced resistance from pupils implementing new seating arrangements (in pairs), issues of gender, ability and lone pupils were evident across most of the teachers.

"...the very first week I had a child turn away from his partner because it was a girl…" Teacher school R.

Two teachers identified pupils who wanted to work on their own, and there was evidence that individual pupils not wanting to work with higher/ lower ability pupils as well as some parings:

"...there are pupils who might not get along well with somebody and just to ensure that the class runs smoothly and for there to be a happy, safe environment, we do ensure that some partners are not put together…." Teacher school N.

".. I think they found it really hard. I remember having one boy asking me if he can … can he be a partner with someone that can-do similar work to him." Teacher school M.

"… they feel quite set in their ways. They had their own seats all year. They liked to stay there. It was what they were comfortable with…". Teacher school R

Teachers expressed resistance from the senior leadership team (SLT), this ranges from no support, SLT not understanding of the process, and thus being counterproductive.

"… the headmaster and the parents were the ones that were fighting against me for it." Teacher school N.

One teacher found that SLT were demanding things be done in the class that did not support the strategies they were implementing,

"Well, there is because the Head Teachers are trying to push this and this and this…. can you write Welsh comments in the children's books. . Why? I'm the lone voice going, "why?" Teacher school R.

Parental resistance was seen as a hurdle, one teacher received many phone calls from parents unhappy with the strategies that were being implemented in the class:

"Their parents would say, "No. I don't want my child to have to sit next to that person." But after I explained to them that it was a system that we had put in place because it had advantages…" Teacher school N.

"Well parents would come up, if they've got siblings and say, "there's no consistency because that child has had so many Dojos and their brother has been really bad at home because he hasn't" and "But we couldn't change golden Dojo because of parents..." Teacher school R.

Parents' discontent with the new classroom arrangements or rewards systems needed reassurance and explanations from teachers to support the strategies this took time out of the teacher's day.

The teachers own resistance mainly came from other strategies that were deployed in the class that were not conducive with parts of the formative assessment project. All teachers commented how certain elements did not integrate with the mathematics programmes that they had in their classrooms. While they suggested that for these lessons' pupils were grouped by ability or worked on their own, they did have to revise the formative assessment strategies to accommodate this programme of work.

"We didn't find that it made any effect, to be honest with you…. And what I saw some children do was ask the high ability kid which one he was doing and then they know, "I can't do that one" …… It was counterproductive. It wasn't them assessing their own work…. So, we just revert back ..." Teacher school M

Two teachers discussed the initial worries about the implementation of the strategies, with one teacher in particular suggesting that there was not enough time in the day to deliver all the strategies and for their classroom it was important to look at other pedagogical strategies.

Context

Two teachers, who teach in areas of deprivation found difficulties in implementation because of the area that they teach in.

"… especially with the behaviour of where our school is..." Teacher school M.

"We're in a disadvantaged area; it's really difficult to get children to achieve the.   This year we might see a change." Teacher school N.

Cohort dynamics were also an issue within the school context:

"...because last year I had quite a difficult class with emotional behaviour there are a lot of issues…" Teacher school M.

"…… but they came up with a reputation." Teacher school R

While demographics of the school were raised, so was physical space, one teacher reported having a class of 30 pupils and this made some of the strategies difficult. Another teacher discussed that the dissemination to other classes was an issue due to the classroom space not supporting the new layouts of desks.

4.3.5 Facilitators

Within the interview's teachers were able to identify mechanisms that were able to support them to implement formative assessment. The key facilitators were behaviour change, SLT support, stakeholder involvement and adaptation.

Behaviour change

Changing practice was evident in most of the teachers interviewed, teachers also identified through reflexive practice that they have changed their attitudes to pupils. This was particularly evident when they discussed pupils choosing the level of task,

> *"I don't put children in a box anymore, there is no stopping any child achieving."* Teacher school L.

> *"...so, they are able to reach a potential that I was not aware of "* Teacher school N.

Reflexive practice was a common theme with teachers realising that they needed to change their mindset towards classroom practices and their own practice:

> *".. completely adopted so I have changed my practice."* Teacher school R.

> *"... and I think now my outlook and life in general has changed because of it. It's an amazing thing to say, but not just in terms of the class but at home and everything I do and everything I aspire to be."* Teacher school L.

SLT support

Support from SLT mainly came in the form of the teachers being able to disseminate the strategies to other classes and school wide adoption. While some teachers were able to trial formative assessment within other departments, some teachers were able to disseminate on a strategic level and integrated formative assessment into policies and school development plans, this was particularly true for feedback and marking strategies.

Stakeholders' involvement

Teachers who presented strategies to parents, pupils and school governors found that this supported the implementation.

> *"...easier to do in the class if the parents are on board."* Teacher school Q.

> *"And they were ... all very positive about the fact that that was happening."* Teacher school P.

> *"...if a child particularly found it difficult to work with others, is having that conversation sometimes with their parents as well."* Teacher school P.

A common thread was that teachers asked the pupils as stakeholders about the new strategies that were being implemented, an example was the elimination of rewards. Teachers discussed the theory around it and gave the pupils the choice to participate:

"And then I had a chat with my classroom, and I explained, "Do you know the course that I've been on? I'm learning as well. Well, let me explain something to you." And I did the story of how adults live, and we don't get anything back. And then they had ... they discussed well, we get paid to come to work, they don't and things. But then we got into the nitty gritty and they agreed that they were to trial eliminating rewards in class..." Teacher school N.

"Well, we discussed it in class first and we said we were going to try and ... you know, we were part of a project, and the evidence suggests ... I told them all about the evidence, what was expected from them, and they all agreed that they wanted to try it." Teacher school L

Teachers also involved pupils in relation to designing different characters to support the delivery of new strategies.

Adaptation

Teachers were able to adapt the strategies to suit the needs of the class, for example for pupils with additional needs either sat in 3's or they were able to have a permanent seat in the classroom so they could be supported by the teaching assistant while still being a 'talk partner'.

4.4 Main findings

From discussions with teachers, two main themes were identified as core features of the FAIP. First, teachers' perceptions of impacts within the classroom were positive and the strategies created a classroom culture that allowed pupils to improve both academically and socially. Teachers were able to detail, encouraging more independent learners, focused pupils on their tasks, focused pupils' learning, understanding of pupils' progress, self-efficacy, improved behaviour, reduced workloads and improved standards.

The second theme centres around the barriers and facilitators the teachers found implementing the strategies into the classroom. Barriers consisted of resistance from stakeholders, and context. Facilitators were behaviour change, SLT support, stakeholder involvement and adaptation.

4.4.1 Comparisons with previous research

Pupils developing more independent learning skills is key to positive educational outcomes (Black & Wiliam, 2005) and links directly with the goals embedded within the Curriculum for Wales (Welsh Government, 2019a). Focusing pupils on their tasks meant that

pupils were engaged it helped them to understand what was being asked of them and how they would go about the task. This finding was echoed in research conducted by Ní Chróinín and Cosgrave (2013).

James (2017) outlined that formative assessment enables teachers to understand where pupils are in the learning. Importantly, this means that teachers can adjust instruction in real time to accommodate the needs of individuals. This echoes Ní Chróinín and Cosgrave's (2013) findings, whereby teachers who were trained in formative assessment strategies were able to understand their pupils better and highlight learning issues for pupils that needed support. While teachers were able to identify where pupils learning was, the formative assessment strategies also helped to identify pupils that needed additional support. Although formative assessment is not a diagnostics tool, it did provide teachers the opportunity to do this.

Leahy and Wiliam (2014) detailed self-efficacy as pupil's engagement with challenges, in contrast to self-esteem which is pupils feeling good about themselves. Self-efficacy is seen as intrinsic to formative assessment as pupils are more likely to be engaged with the task that they are completing (Clarke, 2014). Teachers need to provide the environment where pupils are engaged in a challenge and want to achieve their goals. Self-efficacy within formative assessment has been linked to higher achievement when students are able to self-assess their work (see Panadero et al., 2017).

Behaviour, particularly social behaviour with their peers, was detailed by all the teachers. Pupils were able to form better and wider relationships within the classroom and outside in the playground, and while there is limited research on this topic within formative assessment, Clarke (2014) does suggest that the social aspect of talking partners is 'significant'. Broadly, socially inclusive classrooms are a well-developed field in educational research, and if classrooms are more socially inclusive this can in turn improve the achievement of marginalised students (Gale et al., 2017).

Reducing workloads was evident with all teachers, this was not just because of the feedback policy but a better standard of work with the use of self-assessment/ peer assessment and success criteria. Reducing workloads is a national priority in Wales (Welsh Government, 2018b) formative assessment strategies for the teachers in this study discussed a reduction in workload. The final impact within the classroom was an improvement in the standards of work that they were presented with. While formative assessment is predominantly focused on teachers understanding where pupils are within the learning, thus being able to make changes in the instruction, it has also been synonymous with raising

standards and particularly for low achieving students. Formative assessment is promoted as being able to raise pupils' standards (Black & Wiliam, 1998; Leahy & Wiliam, 2015) and thus an attractive strategy to policy makers and governments.

While teachers discussed that there was resistance from other staff, parents, SLT and pupils, some teachers felt that they had to focus on implementing the FAIP regardless. This is evident in other research, as described by Ratnam-Lim and Tan (2015) who found that the 'rich underlife' and actors that surround school effects classrooms and policies that are being employed within them. Other staff members are not willing to engage in different ways of teaching and note negative perceptions of parents when implementing formative assessment practices in the classroom. This could also be said with regards to the lack of support from SLT through other policy focuses, and Lucas (2020) details teachers who are to be successful must navigate all levels of the school to make changes within their classes. Black and Wiliam (2009) for see that the lack of support for policy change within the management of school will hinder the implementation of formative assessment principles. However, including parents and other staff were also seen as a facilitator to the implementation, Barnyak and McNelly (2009) see that parents can be facilitators and barriers to change within a school setting.

Resistance from pupils took the form of learners who did not wish to sit by certain pupils nor wished to engage in peer review, indicates that these barriers need to be considered within the initial implementation plan. Smith and Godard (2005) found that where pupils were resistant to the strategies, a solution was to improve their understanding of what was happening to them. Involving students as stakeholders within their education is more likely to enable change to happen (Nthontho, 2017). Teachers within this study did find that involving the students in the planning and co constructing the success criteria facilitated changes being disseminated within the classroom. Parental involvement via activities such as reading with pupils and attending progress reviews with the class teacher, actively involved parents to support learning and has been shown to improve outcomes (Hornby & Blackwell, 2018).

Wylie and Lyon (2015) suggest implementation issues can be overcome with support for the teacher, through either more training or being able to reflect on practice. Implementation in schools can be a difficult and complex endeavour, and in this study, teachers were faced with difficulties from other staff this impeded the dissemination school wide. Educational change is discussed at length by Palumbo and Manna (2019), schools are organisations that want to maintain the status quo, tensions are created when there are

different 'values and cultures' and thus make implementing new policies difficult. Guidance written by EEF (Sharples et al., 2018) suggests that for implementation school leaders need to be the driver of successes supporting and monitoring staff. Teachers in this research indicated that support from school leaders was a facilitator to being able to implement the strategies. Adaptations are also important for successful implementation (Sharples et al., 2018) and the teachers in this research were able to adapt the implementation to suit the needs of their classroom.

Finally, teachers demonstrated reflexivity within their practice, and teachers' ability to reflect can improve their practice (Stingu, 2012). While some teachers were able to reflect that they were putting pupils in boxes—with regards to their standardised levels—engaging in this formative assessment project allowed them to reflect on this and ultimately change their practice.

4.5 Study 3: Focus groups with pupils

"And it's better because it's like you've got a teacher beside you"... Pupils school N

4.51 Research focus and question

While the evaluation wants to understand the impacts of the formative assessment teacher professional development programme, it was imperative that pupils' opinions and feedback was evaluated to understand class level impacts and perceptions of the FAIP. Although quantitative data was collected (see Chapter 5) using a mixed methods approach can support the contextualised evaluation of the programme which is important for Theory of Change research (Blamey & Mackenzie, 2007). There is little research that explores formative assessment from the perspective of the pupils (Cowie, 2018). While there has been a general consensus that children as participants are unreliable, children have the capability to validate their own world just as adults do, (Kirk, 2007). They are also able to share experiences and insight of formative assessment strategies in the classroom, and this is an important part of this research (Cowie, 2018).

4.5.2 Research question Study 3 aims to answer the following research question:

What are pupils' experiences and perceptions of using a range of formative assessment strategies?

4.5.3 Participants

In total 57 pupils from eight schools participated in the focus groups in both Welsh and English medium schools. Informed consent was gained from the pupils' parents' or caregiver. Pupils were chosen at random by the class teacher. Pupil assent was also gained before the focus groups went ahead. The pupils age ranged from 8 to11 Years (in Year 4, 5 or 6). Table 4.1 for the pupil characteristics and Table 5.2 details the schools' characteristics.

Procedure

Discussions took place with the class teacher prior to the focus group to clarify the strategies that had been focused on, and this allowed the researcher to finalise the focus group questions. For example, if the class teacher implemented talk partners, pupils would be asked *"can you explain what a talk partner is?"* followed by *"how did you find this new way of working?"* additional prompts would be used *"do you like this way of working?"* or *"did you find this helped with your work?"*. Questions were open ended and tried to be as non-

directive as possible so that the pupils could respond truthfully and not what they may perceive as socially acceptable answers (Bell, 2007). Each focus group comprised of between six and eight participants and lasted between 15 to 40 minutes. Before each focus group discussion commenced, the researcher clarified with the pupils the purpose of the research and co-constructed the ground rules[16]. To try to put the pupils at ease they were shown the recording equipment and how it worked. Pupils were informed that they could leave at any time should they no longer want to take part. Questions were designed to give an insight to the understanding of pupils' experiences of using the FAIP strategies in class; provide information to help evaluate the outcomes pupils achieve; and provide time for pupils to freely discuss their experiences.

Analysis

A thematic analysis approach was employed to evaluate the findings (see Chapter 2 for a discussion on the process used). The data was coded into the different strategies (e.g., talk partners and growth mindset), and questions centred on pupils' understandings of each strategy and the outcomes of the strategies.

4.6 Results

The themes identified from the thematic analysis are shown in Table 4.3 and were as follows: talk partners; mixed ability grouping, including differentiation of task; elimination of comparative rewards; growth mindset; feedback; and meta-cognition. Each theme had a sub theme understanding and outcomes, some elements pupils were able to discuss in further detail, this included some difficulties they experienced.

Table 4. 3 Themes from the focus groups.

Theme	Talk partners	Mixed ability	Elimination of rewards	Growth mindset	Feedback	Meta-cognition
Sub themes	Understanding	Understanding	Understanding	Understanding	Understanding	Understanding
	Outcomes	Outcomes	Outcomes	Outcomes	Outcomes	Outcomes

---

[16] Pupils co constructed and agreed to the following rules: respect each other opinion, do not talk over people, one person to talk at a time, hand up if they want to say something.

	Overcoming hesitation	Difficulties				

Talk Partners

Understanding

Pupils were able to explain the process of how talk partners worked within the classroom, that the allocation was randomly done by the teacher each week, Clarke (2014) discusses the many benefits of having partners that change to enable pupils to have a 'rich diet'(p,57) of talk partners.

> *"Our teacher picks out of a pot who is your partner and we're with that partner for a week and then we change again."* Pupil School O

> *"So, even if you don't like them or you don't play with them, you still have to kind of learn to deal with them and work with them…"* Pupil School L

Pupils were able to demonstrate the purpose of having a talk partner and discuss some of the necessary components that make the interaction productive Clarke (2014) discusses how to introduce this strategy and the successful components needed to be a 'good talk partner'. Talk partners are a way for pupils to become learning resources for one another (Leahy & Wiliam, 2014) amongst other important purposes discussed later:

> *"If you want help with you learning partner[17], and they try and help you, you can't just be silly…"* Pupil School O.

> *"You have to talk in a nice way, and you have to really like talk really sensible when you're talking because you're not allowed to do stuff that you're not supposed to say, or something …"* Pupil School N.

> *"When the teacher asks you to discuss something, you can have help with them, and you have to discuss with each other and find new ideas."* Pupils School O

Outcomes

---

[17] Learning Partners or discussion partners was used by some pupils to describe talk partners.

Pupils understood what a talk partner was, they were able to see that the strategy had other impacts, such as academic support and social relationships. Most of the pupils from each focus group discussed the positive impact on how they were able to socially integrate with a broader range of pupils:

"Cos at first, I didn't have any friends, but after…well, I did have friends, but after we started this, I got more of them…" Pupil School M

"You kind of get to know them more, cos like…you just like…you don't really play with them, cos you like different things, but if you're discussion partners, you might have to try and get to know them…You might think better of them" Pupil School L

"…yeah, you can meet new people and become friends with them." Pupil School R

"And then at the end of the yard time, normally, everyone's playing happily together." Pupil School N

"You can get nervous, like if it was your friend then you wouldn't be nervous, but if it's like somebody you don't like your friends with, then you're a little bit nervous and then you'll make friends with them." Pupils School P

"It makes me feel more popular." Pupil School M

Pupils were able to identify how there talk partner was an academic support, and thus supported their learning journey. There were also able to discuss how this has improved the work that they are producing.

"The reason for why I think talk partners is a good thing is because let's say you're stuck on something; you can just ask them. "Can you please help me with this?" Pupil School R

"Yeah. Learning partner is you can learn different things with them if you don't know one thing, the learning partner would know another thing you don't know." Pupils School P

"It's improved my work because if I'm stuck, I normally ask the Teacher, but if I've got a talking partner, and I can just ask them…" Pupil School N

"My work has improved now because I haven't been talking about like my favourite game or something…" Pupil School N

Difficulties

Pupils discussed that one of the issues that they have with talk partners is copying, while some pupils did not identify it as copying, they identified that they would become annoyed at people looking at their work as well as the pairing being one sided with the workload being carried by only one of the pupils.

> *"…but sometimes partners just want to copy you and don't want to be with somebody their level so they can't copy them, so they pick the same paper as you and they can't do it, so you have to help them a lot."* Pupil School O.

> *"Because you are there to help them, if they look at your work it's not really cheating, because you would have just told them otherwise. But it does just get a bit frustrating to have someone's head just go …"* and *"No. We're quite aware of people looking at our work in a sneaky way…"* Pupil School R.

> *"Sometimes they copy."* And *"When you have to do all of the work."* And *"they're just like, "what's that, what's that" and then their whole entire work was not helping, just me."* Pupil School T.

> *"I would just say, "if you want, I can just help you with the answer, but I would prefer it if you don't actually look at my work."* Pupils School Q

Pupils were also able to articulate the frustration that they sometimes feel with having talk partners, this came with certain pairings.

> *"Sometimes they annoy you and when you're trying to work, they disturb you and try to talk with you all the time about stupid things and stuff."* And *"If you're trying to do work as well, they're gonna like be silly and just bump into you so you mess up all your work and you're gonna have to do it again."* Pupil School M

> *"Cos he might be a bit annoying and kind of say… I'm trying to get on with my work and he's like, "oh but, I've seen this new game" or something and you just kind of…you just want to concentrate on your own work"* Pupil School L

> *"It's because maybe…like maybe they bully you sometimes, so that could make them annoying as well..."* Pupil School M

> *"Because like if you talk too much or stuff like that or if you constantly ask them for help, and they get annoyed, they would just move away from you…."* Pupil School N

Overcoming hesitation

Learning to work in a different way

Some of the pupils detailed how they had to overcome hesitation to the strategies, they were worried about sitting with a different gender or wanting to be with pupils they had already established friends with.

> *"I felt kind of…I wouldn't say really shocked, but it was kind of like a new step for me."* And *"...wouldn't sit by the boys, but now like, I sit by the boys for the partners."* And *"cos sometimes I am nervous about girls, when I was in year 4, cos I always liked to sit by my friends."* Pupil School N

> *"At the start I was thinking, because some of the boys are a bit silly in class, I was thinking that they would distract me. But they didn't."* Pupil School Q

Mixed ability grouping and differentiation of task

While talk partners were used for pupils being resources for each other it also allows for mixability grouping to happen naturally within the classroom. Ability grouping (grouping, setting or streaming) has been used in the British education system for decades, this is where pupils are grouped with pupils the same level of current achievement and have tailored instruction depending on high, middle or low ability. Research suggests that lower ability groups are more densely populated with disadvantage students and there is lack of mobility from lower to higher groups (Francis et, al., 2019). Concurrently research also suggest that ability grouping has a relatively small effect for high ability pupils (Ireson et al., 2016). Mixed ability teaching has seen teachers tailor instruction and tasks to suit a wide range of abilities. Differentiation of task within this programme mostly centred on teachers designing a task with a different level of difficulty and the pupils could choose the task that they believe is suited to where they are in the learning journey, within the sample of pupils the differentiation was used with the aid of the 'learning zone' diagram, this would help the pupils and the teacher decide if they needed to choose a different task or needed further instruction to be able to complete the task based on the perceived level of difficulty. The learning zone is a way for pupils to assess where they are in the learning and give teachers a way to formatively assess and provide feedback (Hattie & Clarke, 2019).

Understanding

Pupils were aware what 'ability group' they were in and if they were sitting with a higher or lower ability pupil and how this could support each other's learning:

> *"It could help you or they could…maybe you can help them."* Pupil School M

> *"So, if you're at the lower level and you're not quite sure what to do, without the discussion partners, you might just go and sit with someone at the same level as you when you wouldn't really learn more, but with the discussion partners, if they've got a higher ability than you, then they can help you and then you can be at a balanced level."* Pupil School L

> *"...if you've got a talk partner who's a higher level, it actually brings your learning up…"* and *"Yeah, and for the higher level, it brings say like talking skills up…".* Pupil School L

Differentiation of task

Pupils were able to understand and describe the element of differentiation of tasks.

> *"I choose like if I want to choose easy, hard or medium depending on how hard the work is and how I feel. So, sometimes I might choose the easy one, sometimes I might choose the hard one."* Pupil School R.

> *"So, we have three different tasks, and each task has three different levels of hardness. Yeah, difficulty. Yeah, and so you get to choose which tasks you want to do first and which level of difficulty you want to do. If you get halfway through and you're finding it really difficult, we have challenge cards. So, there's a red one when you're really finding it difficult. Green.", "A green one when it's a bit too easy and orange one where you're being challenged."* Pupil School L

Difficulties

While the pupils could see that being with a different level of pupil could support learning there were incidences where pupils explained that there were some negative consequences for both the higher and lower ability pupils.

> *"Sometimes if they ask you a question and they're like a lower level than you, you're kind of just like…and they ask loads of questions, you're kind of like, are you actually making me do all of your work, or actually, do you need help."* Pupil School O

"Well, holding each other back I can do more work, but they're stopping me.." Pupil School R

"Sometimes it can feel like someone on a different level than you and you're saying, like, please can I copy all of your work." And *"Yeah, and just being lazy."* Pupil School O

While some pupils explained that they felt pressure and had negative feelings:

"... sometimes, you feel very pressured to do the exact same amount and you don't really know everything that well, but then they do. So, they're rushing ahead, and you feel really pressured and sometimes you feel, like, stupid." Pupil School R

"...and it makes me feel like I'm near the bottom all the time and I can't do anything." Pupil School P

Outcomes

The pupils were able to detail that while there were difficulties, these strategies allowed them to be focused on their work and for them to understand where they were within their own learning journey and for the teachers to know as well.

"A green one when it's a bit too easy and orange one where you're being challenged. So, if you're on the red card, then…A teacher will come and help you…" Pupil School Q

"Yeah, a teacher will come and help you or if you still don't understand it, you can change to a lower task." Pupil School L

Elimination of comparative rewards

Elimination of comparative rewards was a strategy only discussed with two focus groups, and only two of the teachers in the sample trialled it. The elimination of rewards (e.g., stickers and prizes) within this formative assessment project was to remove external rewards in favour of feedback. External rewards can undermine future motivation, and children become focused on the reward and not the task (Sears & Pai, 2012). Black and Wiliam (1998) discuss that if a classroom culture is dominated by rewards pupils focus on the best way to get the reward rather than the learning. Rewards systems have limitations; for example, teachers can use them inappropriately and pupils' intrinsic motivation is not improved; and pupils can take a narrow view of tasks and can become negative (Ilegbusi,

2013). Pupils who do not achieve will internalise negative self-belief. Feedback should be focused on specific learning objectives and should suggest specific areas of improvement including how to improve the quality of their work (Clarke, 2014).

Understanding

Pupils in both focus groups discussed how the rewards were being used as bribes for them to do work and monitor behaviour.

> *"It's a bribe because like you say, "oh, I didn't get that this week, I'm not gonna get it", or you say, "oh, I'm gonna try my hardest to get that", so you're not trying your best actually."* Pupil School N

> *"Because kids only try if they want to get rewarded"* Pupils School N

> *"It's a bribing game because like with those naughty children where miss [teacher] will bribe them, saying, like, "well, if you be good, you might get the band[18]"* Pupil School N.

Pupils were able to detail that the reward system, would elicit negative feelings within themselves, and that the system was unfair.

> *"….in one week, like one week, you might feel I've tried really hard, yeah, and you're at the lower level and the higher level would be…they haven't tried as hard, but they've done a bit better work. They might get it and you would feel, "but I tried my hardest".* Pupils School L

> *"Yeah, you might feel like even if you've tried your hardest. Say, even if you did really good work, yeah, but like you didn't get star of the week, you'd feel "oh, I might just be bad and I might just think I'm good", but really you tried your hardest".* Pupils School L

> *"..so, we're saying there used to be a couple of naughty people in our class, then a Teacher would go, "oh, if you be good this week, you might get a band, and they were like "oh yeah, I'll be good then", but then for the people whose good all the time, but we don't get it".* Pupil School N

---

[18] Pupils could choose a rubber bracelet as a reward, the pupils called it 'getting band'.

Pupils were able to detail that the focus was not on their learning, but that the motivation was to get the reward.

"Another reason we stopped is because kids only try to get the band and then they just stopped after they got the band. They wouldn't work hard." Pupil School N

"Because kids only try if they want to get rewarded.". Pupil School N

"...because they weren't getting anything and they'd feel like they were missing out and I guess the same people might get it over and over again and you'd think, "oh, well, I don't need to work, because I'm not gonna get a reward anyway". Pupil School L

Outcomes

The pupils understood that without the reward system they would focus on the work they were doing, get feedback in other ways to support their learning and that removing the reward made the classroom more equal.

"But without having rewards, it's easier to...like you want to work harder as you're not getting a reward, nobody is, and everybody's being treated the same." Pupil School L

"Yeah, getting rewarded by someone telling you you've done well is better than someone giving you something, like just a colouring pad or something, cos you probably won't keep it, it won't be there for long, but if you had someone tell you you've done good work it's better..." Pupil School L

"On Friday, because we don't do it anymore, I think I've learnt more...in my opinion, I think my brain is growing ...." Pupil School N

"When we get praise now, it's like "well done, well done" and they message your Mam's.". Pupil School N

Pupil: *And also how we get like told that we've done good in our work, we also get a high five from the Teacher.*

Researcher: *Yeah. And does that feel good?*

Pupil: *Yeah.* Pupils school N

Growth Mindset

The work of Carol Dweck has brought growth mindset theory and practice into mainstream focus in many classrooms. The neuroplasticity theory that underpins growth mindset describes how peoples' attitudes to intelligence can affect their achievement (Yeager et al., 2019). When people believe in natural abilities and talents 'fixed mindset' they are less likely to think that they can achieve. Whereas a 'growth mindset' is a person's ability to see *"... abilities can be cultivated and developed through application and instruction."* (Dweck, 2009, p. 2). In schools that promote Growth Mindset as an intervention or approach, pupils have lessons on how the brain is malleable and that intelligence is developed not innate. Many of the schools in this FIAP study taught growth mindset using the idea of a green and red brains. The red brain represents a fixed mindset, where intelligence is seen as an innate ability, and when faced with failure pupils can avoid challenges, give up on tasks easily and does not see the purpose of efforts. The green brain the growth mindset brain, understands that effort and challenge will support completion of the task (Clarke,2014). Praise linked to intelligence is replaced with praise crafted around effort. However, the effect on pupils' achievement is mixed. Li and Bates (2019) comprehensive study found no effect on standardised test scores following an evaluation of a growth mindset programme in schools. There is also only a very limited evidence base in the published literature for the positive effects of growth mindset on learner outcomes. However, Rege et al. (2020) argue that this is not a true representation of effect sizes in real world settings, and in a large-scale research project they found positive impacts on achievement and in the field of psychosocial interventions growth mindset seen gains on depression and anxiety (Schleider & Weisz, 2018).

Understanding

> *"Growth mindset is like when you…if you can't do something, you say, "I can't do it yet", but if you have like a red mind, and you will just be like, "oh, I can't do this at all, I'm giving up, I can't do this".* Pupil School O

> *"...there's a different way of looking at things. If you get, say, something wrong in your work … instead of saying, "oh, I did it wrong", say, "doing things wrong helps me get better".* Pupil School Q

> *"Growth mindset, so, say five plus two is seven, and I put it down in my book at like nine. If I had a growth mindset, I wouldn't go off crying and saying, "oh, I got this wrong", you would try and think and put the right answer down…"* Pupil School L

"It's kind of ... it's a different way of thinking about your work if you can't do something. It's like instead of being frustrated and giving up, you would kind of just have to look at it in a different way. Like just instead of, say, looking at the final thing you're trying to get, to maybe like making small steps." Pupil School Q

Outcomes

Pupils in the focus groups were able to identify how having this strategy implemented in the classroom impacted on their work, self-efficacy, more positive and were engaged with their work.

"It makes you calm and focus on your work." Pupil School P

"It helps ... like for us to not feel negative all the time." Pupil School Q

"You have to be positive and if someone says they can't do it, you have to say, "yes". Pupil School P

"...before we started doing growth mindset, I was a bit like, "I don't really want to do this anymore, it's getting a little bit boring", but with the growth mindset, I've carried on doing it and it's nicer to do it now." Pupils School L

"...And embracing challenges." Pupil School L

Success criteria

Success criteria is a set of steps that a pupil will be able to follow in order to complete a task that support the learning intention. This process supports the pupils to break down the task into small steps to help them learn, success criteria also can be used by the pupils to identify the parts of work that are important and can improve the quality of work that they produce (Leahy & William, 2015). The learning intention is the skill that is being taught, this can be a closed skill for the task or an open one where the pupils will be able to transfer to other areas an example would be writing a letter to a newspaper, the learning intention would be writing a letter and the success criteria would incorporate date, address, appropriate opening and closing of a formal letter Clarke (2014).

Understanding

"It involves, say if you was doing maths, and it says, "L I to solve problems", but the success criteria is like you need to follow it to do good work." Pupil School P

"And you just have to follow the success criteria … you have to check while you're doing your work.". Pupil School R

Outcomes

Pupils were able to detail that the success criteria made the task they were completing more strait forward. They were able to check and improve the task that they were completing.

"...has helped because then you can make sure your keeping to the right thing and don't have to keep looking back at it." Pupil school O

"It makes the work a bit more straight forward, Cos when you look at the success criteria when you're working, then it like gives you more to think about it and then more to think about the work." Pupil School P

"Like when you said if we check our work … if we show after we've done all our work, so if it was English comprehension, after we had showed the teacher she might say, "oh, you forgot this bit of the success criteria. See if you can fill it in anywhere", and … yeah. Yeah, I think it improves our work and then we know for next time if we're doing a similar type of work … for next time that we should remember that."* Pupil school R

Success criteria being co-constructed allows pupils to have ownership of their own learning, and become more independent (see Clarke, 2014 p 86 for a full discussion). The sample within this focus group were able to detail that they were able to co-construct the success criteria.

"Before we do the work, then the Teacher…then we'll put our hand up suggestions that we need to do in the task, and he writes it down and then he uses that success criteria for our work." Pupil School O

"So, when you do your success criteria, you don't try and like do a capital letter, cos like you should do that if you're year 5 and 6, like you've tried to do what would challenge you…" Pupil School L

While some pupils discussed the success criteria was co-constructed some pupils had a different experience.

"And I don't think I really need to be reminded of capital letters." Pupil School R

"Sometimes she asks us what we think, and we should do in our success criteria. So, then we do our ideas and then she writes it down. But most of the time she just chooses the important stuff like capital letters, full stops, punctuation, if it's in English" Pupil School R

Feedback

The provision of effective feedback to learners is an important subject area in education (Clarke & Hattie, 2018). Within this study, teachers were asked to experiment with feedback at different stages of the lessons, including peer and self-assessment. Within this sample of pupils, feedback was commonly expressed through the use of colour coded pens for visible feedback, mid-lesson stops to allow teacher input and/or pupil discussions, and self- and peer feedback and the use of the learning zones[19]

Visible feedback:

"So, say you write something like a story, anything like that, and you…the pink one is what you've done well, and green is what you can try and do better next time." Pupil school O

"…the Teacher will take you out of the lessons and things just to like go over your piece of work and if you've done something well, he'll tell you what you've done well and he'll like highlight it on the success criteria, which is a list of things that you have to do and he'll highlight it pink and then if you need to do something better, he'll highlight it green and then he'll tell you to re-do it and he'll tell you what to re-do and stuff." Pupil School O

"Sometimes, Mr [Class Teacher] lets our talking partners view our work and then get a red pen and if there's a capital in the middle of a sentence or something, they put a circle around it and then they give the book back when they're ready to the person that the book owns, and then they look that if there's a red circle and then they have to get a purple pen and do it correct." Pupil school L

Peer assessment

---

[19] A visual aid for pupils to identify different zones of learning (from the centre: comfort zone, learning zone and panic zone).

"So, sometimes going back to the talking partners, the Teacher tells you to put each other's books on top of each other and go through the work with your partner." Pupil School O

"...then you get your books with your talk partner and then you get a purple pen and we put our books on top of each other and then, … while I was reading it or we're both reading it, if we spot a mistake she writes in it with her purple pen." Pupils School R

Mid lesson learning stops

And like he'll stop you, like we have a certain amount of time to do it and he'll stop you like halfway or something, and he tells you to check if you've done everything before you carry on." Pupil School O

"...she usually just does this thing called pit stop. Like, before we start our work, she says, "okay, in 15 or 20 minutes we're going to have a pit stop" Pupil school R

Learning zone diagram

"You have to be in your…you're not allowed to be in your comfort zone or your danger zone, you have to be in a stretch zone." Pupil School P

Outcomes

The strategies above allowed the pupils to see where they were within their own learning, where there were mistakes and allowed them to be resources for each other. The teachers were able to identify mistakes and know where the pupils were within the learning journey.

"A green one when it's a bit too easy and orange one where you're being challenged. So, if you're on the red card, then… a teacher will come and help you or if you still don't understand it, you can change to a lower task." Pupil School L

"So, while … when our teacher is looking at our work, rather than using our normal pen, she can spot it easily with our purple pen. So, she can spot the mistake." Pupil School R

Meta-cognition

Meta cognitions typified definition is 'thinking about thinking'. Within a school setting pupil are given a set of strategies to explicitly think about their learning (Martinez, 2006), Self-regulated learning can be broken into three areas: cognition, meta cognition and motivation. Pupils will identify their own strengths and weaknesses and be able to plan, monitor and evaluate their work (Muijs & Bokhove, 2020). Teachers have introduced certain characters that are attached to tools that pupils need to understand and discuss learning, for example 'Concentration caterpillar' (Clarke, 2014). Giving pupils a shared language, the ability to think about their learning and take responsibility for their own learning and identify which strategy they need to complete tasks. Pupils were able to articulate their understanding of meta-cognition and how this impacted the tasks.

Understanding

"It's like animals, they show like learning .... but what you need to know if you need to learn." Pupil School P

"It's when you like concentrate, then you're like the concentrating caterpillar and like you can think of it and then if you're like helping others, like if they're stuck, then you're a teamwork tiger. So, then you help." Pupil School P

Outcomes

"...so, I couldn't concentrate. So, I went to the round table, and I thought, "oh, I got that wrong, let me do that again", and I actually improved my work." Pupils School N

"When you've done something, you can think, "what learning powers have I used?". Could I use more? Could I have used, like, more, could I have done a better story by concentrating more." Pupil School L

4.7 Summary of findings:

The purpose of this research was to look at pupils' perceptions and experiences of having formative assessment strategies in their classroom. Pupils were able to describe and understand each of the strategies that were implemented.

4.7 1 Comparisons with other literature

Pupils discussed that they were able to have relationships with other children outside of their friendship groups, and they felt this culminated in a more socially inclusive classroom. Some individual pupils were able to explain that they had made more social contacts in the classroom, and this feature has been shown in other studies to improve pupils' mental health and academic achievement (Ng-Knight et al., 2019). Socially inclusive classrooms benefit pupils in school and later in their working life as they learn to work with different people (Hymel & Katz, 2019).

However, there were instances where pupils were not at ease with talk partners and discuss the initial hesitation at different gender, abilities and copying work. Clarke (2014) suggests the teachers need to create the right classroom culture for the pupils to be able to benefit from talk partners. While cheating has been a concern in educational literature, particularly in higher education, formative assessment practices are championed as a way to combat cheating (Trail-Constant, 2019).

4.7.2 Mixed ability

The pupils discussed they were able to support each other and had different skills to bring. Other research on mixed ability grouping suggests improved educational outcomes for pupils (Francis et al., 2017). However, there were cases where higher ability pupils saw the mixing of ability as a negative strategy. Research has suggested that there becomes a hierarchy between the students, and there are mixed results when looking at whether mixed ability grouping mitigates the attainment gap (McGillicuddy & Devine, 2018). Some pupils also detailed that they felt pressure to keep up or felt negative being with a higher ability and this reinforced their low ability status. This reinforcing could have a negative effect on their attainment and self-worth as pupils can internalise their status (Snell & Lefstein, 2018) and should be explored in more depth.

4.7.3 Elimination of rewards

Pupils explained rewards would lead them to internalise negative feelings, and that rewards were given to bribe them to do work and manage behaviour. Pupils were able to understand they were not motivated to learn through the use of rewards, and in this case the use of rewards appears to only tap into the external motivation. This has also been identified by Leahy and Wiliam (2014) and Ryan and Deci (2020). External motivation through reward systems reduces effectiveness overtime, and "… distracts students from true independent learning." (Adamma et al., 2018, p. 53). The pupils from the two schools who trialled this

strategy discussed they were able to focus on work and teachers using praise was rewarding to them. Ryan and Deci (2020) suggest that a deeper understanding to the theories of motivation in education needs further research, but that intrinsic motivation can support pupils' wellbeing. This is important area for the Welsh Government and the Curriculum for Wales, where health and wellbeing is one of the core Areas of Learning and Experiences (Welsh Government, 2019a).

4.7.4 Growth mindset

Growth mindset allowed the pupils to focus on the positives. They discussed enjoying working within this positive mental mindset, and that they were not disappointed if they made mistakes. Embracing challenges is important for pupils and can improve their outcomes (Ng, 2018). Research also suggests that there is a link between growth mindset and intrinsic motivation, which culminates in increased more motivation, and self-regulation. These are all imperative for a pupil's success in education (Ng, 2018).

4.7.5 Success criteria

Pupils discussed the success criteria gave them the building blocks to complete the task, see where improvements were needed and were actively thinking about the task, thus engaged in their learning. A small number of pupils were despondent about the use of the success criteria. This was also identified in research from Crichton and McDaid (2016), and in this research the pupils discussed the success criteria was not always co-constructed. Clarke (2014) emphasises the importance of pupils being involved in creating the success criteria to be more effective and tailored to the pupils learning.

4.7.6 Feedback

While the theme of feedback was large and multifaceted, pupils suggest they were able to use feedback to improve their work, identify mistakes and, most importantly, identify how to improve their work. Pupils discussed the use of highlighters or different colour pens. Clarke (2014) terms this 'the polishing pen', and each colour is synonymous with improvement areas or areas that are good. This type of visible feedback can help pupils identify areas they need to improve and allows them to be encouraged by the positive feedback (Wisniewski et al., 2020).

4.7.7 Meta cognition

Meta cognition strategies have been seen to improve pupils' outcomes, wellbeing, agency in the classroom and have been linked with educational policy in England (Perry et al., 2019). While, only on a surface level the pupils were able to demonstrate, understanding, they could improve their work and were engaged in thinking about how their work, the meta cognitive strategies gave them the language to do this. Recent guidance from the EEF suggests that pupils can progress seven months if the strategies are used correctly in the classroom, and that pupils can also improve their self-regulation skills to help their learning (Quigley et al., 2018).

4.8 Discussion for Study 2 and 3
4.8.1 Strengths and Limitations

There has been a growth in participatory research with children, and learners in schools are now seen as social actors that can construct their own world. Another strength to this research is that it gave a voice to the pupils in research can sometimes see pupils as passive actors. However, engaging learners in research can help us to understand how taking a look at their world can articulate impacts and difficulties that can help to navigate implementing interventions and how "educational practices are encountered and conceptualized by learners" (Todd & Nind, 2011, p. 116), and helps us to understand the nuances that can accrue when interventions are deployed in the classroom (Crichton & McDaid, 2016).

While qualitative data can only yield perceptions and experiences of a phenomena, it has allowed us to identify some of the formative assessment strategies have been used in schools. In particular, investigating pupil's perception and experiences allows researchers to see how the intervention is received on the people it is intended to benefit, which has been neglected in educational research (Yoon & Templeton, 2019).

Most formative assessment research is designed around tightly organised teaching and learning communities (Wiliam, 2020), or designed around a specific tasks or extended learning activities where teachers receive support from a research team (Hendrich et al., 2016). It is highly likely that there was a wide variation in the way the teachers implemented the FAIP strategies. This research team had no control over the dissemination and implementation of information shared with teachers. The teachers were given a choice and could be flexible in what they chose to disseminate within their classrooms.

Efforts were made to ask the questions in a non-directive manner, so that pupils felt little pressure to answer in a way they thought was acceptable or pleasing to the researcher.

Given the universal approach and multiple strategies trailed in the classroom, at times it was difficult to get the richness and depth of understanding about the implementation of specific aspects of the FAIP programme. Each strategy implemented could be considered fields within their own right, and different ways in which effectiveness and implementation can be measured. Intrinsic motivation and growth mindset need further investigation in neuropsychology and yet there is little understanding how the brain is affected (Ng, 2018). Feedback was another example that could have been explored further, but this was not possible due to the resources and time limitations of this study.

The sample of students was small, and thus could not be representative of all Tier two pupils. As such the sample of teachers was also small. Given the range and scale of the project, representing all learners across the age span would have been difficult, time consuming and required a more substantial research budget.

It would be helpful, therefore, to collect more longitudinal data, given the evaluation window was only one academic year.

4.9 Conclusions to Study 2 and 3

The main aim for the qualitative work in these two studies was for them to be incorporated into the larger evaluative framework to understand the impact of formative assessment described in Chapter 5. Both the teachers and pupils perceive positive impacts of the FAIP strategies. While there were some difficulties for teachers facing resistance, and pupils having to get used to working in a more cooperative way and sharing there learning journey with other pupils. Both data points to a positive impact of the FAIP programme on teachers' skills and learners' experiences in the classroom.

This research aimed to understand pupil's perception and experiences of universally targeted formative assessment implementation. Pupils were able to discuss positive and informative results, and they demonstrated a good understanding of the key features of effective formative assessment strategies to help them learn. This research demonstrates that a unique train-the-trainer intervention, with autonomy given to classroom teachers, can have a positive impact on the quality of provision in schools. Using pupils as resources for each other (talk partners) had positive effect on the social world of the pupils this could also be investigated in more depth. Giving a voice to pupils can demonstrate to policy makers that

the classroom is more than test scores they can be drivers of change, and in a wider sense support decision making as to what happens in *their* classrooms (Nelson, 2015).

4.9.1 Policy implications

The 12 pedagogical principals in the new curriculum can be directly linked to the strategies employed within the FAIP. For example, pedological principle 7 "means employing assessment for learning principles." (Welsh Government, 2019a, p. 50) and, more broadly, assessment and classroom culture should encompass understanding the learner so to make changes to teaching or provision.

Teachers who were trained were able to disseminate to the classroom and that pupils understood and discussed the impact on their learning. In this context the FAIP is a suitable programme to help schools develop effective teaching strategies in line with Curriculum for Wales.

While most formative assessment research is tightly controlled, teachers in this project were able to use the strategies and pupils understood why they were being used; the FAIP training successfully enabled teachers to move beyond the understanding of abstract concepts and towards a practical method for to teachers the disseminate formative assessment in the class (Yoon & Templeton, 2019). Universal approaches interventions such as this that target entire pupil populations have been overlooked by policy makers. There is some evidence that this type of project can have positive impacts and supports the discourse on the of universally targeted interventions (Greenberg & Abenavoli, 2017, p. 44).

Reducing teachers' workloads is an important focus for the Welsh government (2018b), and while this research is only a small part of the conceptual landscape there is some evidence that using some formative assessment strategies for the teachers in this study supported reducing their workload.

Chapter 5

"The heroic role for the economist is to do the solid research that provides the correct and reliable information to the policymaker that tightens the policy maker's priorities around correct parameters, and this leads to optimal policies." (Pritchett, 2009, p. 42)

5.1 Cost Consequence Analysis of FAIP

5.2 Abstract

Given the limited use of economic evaluations in the education sector, the 'North Wales Formative Assessment Project' provided an opportunity to demonstrate the utility of using health economics methods to evaluate a regional school improvement project. This evaluation will utilise a Cost Consequence Analysis (CCA) approach to provide an economic analysis to support decision makers. This study used a mixed method approach with a matched control group as part of this CCA approach. The results indicate that the cost per pupil was £67.55 (inflated to 2023 prices) and no statistically significant outcomes were found across all the measures used (i.e., national test data, CHU-9D, SDQ and QoLS). Qualitative data suggested a positive impact of the FAIP was experienced by teachers and the pupils, including improved standards of behaviour, reduced teacher workloads, and an improvement in the standard of learners' work. Pupils discussed understanding the rationale behind the new formative assessment strategies and reported being more able to focus on the tasks they are given that helped them produce a better standard of work. These findings suggest there needs to be more consideration of the qualitative impacts of outcomes that cannot be monetarised, and further research is needed to standardise costing in the education setting.

5.2.1 Introduction

There is a drive for education systems to be become more evidence informed to help education professionals plan and implement provision that can improve learner outcomes (Gorard et al., 2020). Governments are supporting schools and practitioners to identify evidence informed practice, and evidence repositories such as the Education Endowment Foundation (UK) or What Works Clearing House (USA) now provide schools with important summaries of research evidence and information on more promising programmes (Hilton, 2017). Education has not faced the same scrutiny as other public services and there is a lack of economic evaluations to support decision makers at all levels of the system (Levin &

Belfield, 2015). Simultaneously collating effectiveness and cost data will provide decision makers all the information needed to consider either maintaining or discontinuing with a course of action or programme (Gilead, 2014; Hummel-Rossi & Ashdown, 2002). Policy makers in education aim to maximize the outcomes for learners, and this is particularly important with the cuts to public spending witnessed since 2010. In education there has been significant real-term cuts to education spending over the last 10 years (Cardim-Dias & Sibieta, 2020). The need for greater accountability in the use of public spending has been driven by the 'New Public Management' (NPM) that focuses on public services being managed like the private sector with a clear focus on value for money, increasing efficiency with set targets and monitoring (Talbot, 2009). While the uptake for economic evaluation in education has been slow, there is a small and growing base for this area of evaluation (Levin et al. 2017; Gilead, 2014). One way to show value for money in public spending is through economic evaluations, and these have already been used in health, transport, defence, and the environmental sectors (Cubi-Molla et al., 2021; Gilead, 2014). The health sector provides an example of established economic evaluations that can be utilised for other policy areas (Cubi-Molla et al., 2021).

5.2.2   Method of economic evaluation

The main methods for health economics evaluations discussed in Chapter 1 include: cost-effectiveness analysis; cost-utility analysis; cost-benefit analysis; cost-minimisation; and cost-consequence analysis.

5.2.3   Rational for CCA

The rational for using CCA is described in detail in chapter 2.

Research questions:
- What is the cost of implementing the 'North Wales Formative Assessment Project'?
- What impact did formative assessment have on pupil's attainment, health related quality of life, wellbeing, and quality of school life?
- What perceived impacts did teachers and pupils have of the formative assessment strategies?

5.3 Methods

The research design was quasi-experimental. Quasi-experimental designs work when an event is happening in the real world that would happen independently of the research project, (Thyer, 2012). This is seen in policy roll out or social change (Hakim, 2012). This research design is also helpful for when the research team does not have full control over the disseminations of the intervention. This type of research design has been used extensively in evaluation studies that are large scale. The criticism of this research design is that the researcher has many issues to consider with the internal validity, because of the nature of conducting research in a social setting and the researcher not being able to control for all the variables. Generalisability is affected because of the sample and may not be representative of the population that is under study. The ultimate difficulty in any quasi-experimental design is answering questions when there are so many independent variables. (Eccles et al., 2003). Gray et al. (2005) suggests that quasi-experiment designs work well in an educational setting as the researcher has no control over the formulations of a class they are already formed, particularly in a universally targeted intervention where the sample is large and heterogeneous. Mixed methods were used, to try to capture a wide range of outcomes, the different data collection methods can complement each other and maintain some validity and robustness of the findings (Zohrabi, 2013). Using a mixed methods approach allows the researcher to capture the wider benefits or outcomes, which is important given the complexity of this study.

5.3.1 Intervention

The intervention is discussed in detail in Chapter 2. Teachers were trained in different elements of formative assessment through a train-the-trainer model that helped them integrate formative assessment elements into their everyday teaching. Research was conducted in the schools setting in the classroom or a suitable location that was free from disturbances.

5.3.2 Study population

This study focused on learners in primary school settings in Years 4, 5 and 6 (age 8-11 years) and their class teachers. The schools represented both Welsh and English medium settings and a range of sizes and levels of deprivation (% e-FSM). Table 5.1 details the school characteristics for both intervention and matched control schools.

This population was chosen based on the following rationale: First, in a primary school setting pupils are usually exposed to the same class teacher, and this provides a more consistent dosage of the intervention. Second, the measures employed for this research were

not validated for younger pupils, (under 8 years) and they would have required a parent or teacher to populate the questionnaire for the learner. The Strengths and Difficulties Questionnaire is an example of this. All interviews were conducted in the schools setting either in the classroom or a suitable room where there were limited disruptions. Informed parental consent was gained for the pupils in the intervention arm for quantitative and qualitative data collection. Informed consent was gained for pupils in the control arm for quantitative data collection. Teachers gave informed consent to take part in the interviews.

5.3.3 Recruitment

All primary schools in the second tier of the FAIP were emailed with an invitation to participate in the research study. A purposeful sampling approach was used to identify schools. This involves working with existing networks to identify individuals or group of individuals who are knowledgeable and willing to engage with research (Palinkas et al., 2015). The education provider (GwE) identified 35 schools that have engaged in research in the past and were part of the FAIP. All schools were contacted, ten schools were recruited for the intervention arm of the research. For the control group, a matched sample of schools was identified using school language, size, and percentage of e-FSM as matching criteria. Table 5.1 details the characteristics of the school in both the intervention and control arm (See appendix D1 for control consent). Initially 10 intervention schools and 10 control schools agreed to participate; two schools in the control and one interventions school failed to provide follow up data and were omitted from the study. The intervention arm included nine schools and the control arm included eight schools.

Table 5 1 Characteristics of participating schools

	Schools	Number of pupils of statutory age	Language	Local authority	eFSM %
Intervention	School L	82	Welsh	Anglesey	8.5
	School M	83	Welsh	Gwynedd	19.3
	School N	179	Welsh	Gwynedd	34.6
	School O	87	Welsh	Gwynedd	16.1
	School P	326	English	Wrexham	23
	School Q	57	Welsh	Gwynedd	8.8
	School R	355	English	Flintshire	8.2
	School S	174	Welsh	Gwynedd	3.4

	School T	287	Welsh	Anglesey	29.3
Mean		181			16.8
Control	School A	110	Welsh	Gwynedd	11.8
	School B	118	Welsh	Anglesey	7.6
	School C	57	Welsh	Anglesey	19.3
	School D	308	English	Flintshire	22.6
	School E	214	English	Wrexham	19.2
	School F	55	Welsh	Gwynedd	30.9
	School G	83	English	Conwy	19.3
	School H	112	Welsh	Denbighshire	17.9
Mean		132			18.6

5.3.4 Measures

The measures are discussed in detail in Chapter 2, as reminder the following pre- and post-measures were utilised: (1) The Strengths and Difficulties Questionnaire (SDQ) is a child and adolescents screening tool commonly used to identify behaviour difficulties. The SDQ has been used extensively to assess social and emotional wellbeing as well as behavioural difficulties (White et al., 2013). Formative assessment has been linked to improved pupils' behaviour, self-regulation, and non-cognitive improvements (Dunphy, 2010; Black & Wiliam, 2009); (2) The Child Health Utility-9D (CHU-9D) measures the health-related quality of life and has been validated as a preference weights measure for children (Stevens, 2012); (3) The 'Quality of Life in School Questionnaire' (QoLS) measures the quality of life in school. Some evidence suggests that pupils develop improved relationships with their teachers, enjoy school and feel connected (Ferreira et al., 2014) and (4) National Reading and Numeracy Test data from the pupils in the intervention and control schools. Formative assessment has been linked to improved outcomes for pupils (Black & Wiliam, 2009; Clarke, 2014; Ozan & Kincal, 2018).

Pre intervention measures were administered in the Autumn term of 2018, and post intervention measures were collected in the summer term of 2019. The questionnaires were administered through an online survey tool. The surveys were available in Welsh and English and paper copies were available if needed. National test data was supplied from GwE, it was anonymised at class level limiting any further regression analysis. National test data is

collected at the beginning of the spring term each year. GwE supplied national test data for the academic year 2017 - 2018, to compare and changes in national test data. Parental consent was obtained for the intervention and control arm to collect questionnaire and national test data.

Interviews

Interviews were conducted to elicit the perception of impacts at class level for the teacher and the pupils and discuss any barriers and facilitators to implementation. Interviews were used to collect additional time utilised and costs for teacher to run the project in their school.

Focus group

Focus groups were utilised to explore the wider perceived impacts from the pupils' point of view, this gave the pupil the opportunity to have a voice in the process and as they are the main beneficiaries of the project gaining their lived experience was important. Pupils were chosen at random by the class teacher to participate in the focus groups. All pupils had parental consent and assent was gained before the focus group started.

Observations

Although there are no standardized observation checklists available for observing formative assessment in classroom settings, a checklist document used by Leahy and Wiliam (2014) was used in the current study. GwE school improvement officers were trained on how to use the observation checklist (see appendix A5 for observation checklist). Given the limited control the researcher had over the elements of formative assessment that were disseminated in each classroom, the observations checklist centered on the main principles and the observations were rated on the following scale: Not applicable; Applicable but not observed; Observed but poorly implemented; Observed and reasonably implemented; or, Observed and well implemented. Observations checklists were available in Welsh and English.

Collating costs

Collating costs was from an education provider perspective (GwE). The majority of costs were incurred by GwE, and the main cost was supply costs to cover teachers to attend the training and for staff cost to support the project. GwE have a flat rate for buying out teacher's time for training, £250 for a full day and £125 for a half day. A detailed description on calculating the cost of Business as Usual (BAU) is discussed in detail in chapter 2.

Table 5 2 Cost for teachers' time including on costs.

Cost (Mean)	2018-2019 prices	2020- 2021 prices	2022-23 prices
Teacher cost yearly	£58,544	£60,947	£72,233
Cost per pupil yearly	£3183	£3314	£3928
Cost per hour	£46	£47.89	£56.89

Sensitivity analysis

To test assumptions, sensitivity analysis was conducted on three different parameters of the costing assumptions. Testing assumptions can support robust conclusions (Levin et al., 2018).

The costs were supplied from GwE and are fixed cost and represent the actual cost, as with the other resources such as, translation for training and hiring training venues. To replicate the programme in another area the costs may vary. There are some uncertainties over assumptions made on three main costs.

Sensitivity analysis 1

Average class sizes in Wales are 25 pupils, this study assumed for each teacher that was trained 25 pupils would be exposed to the intervention. Thus, sensitivity analysis was conducted if the average class size was 20 and 30 pupils.

Sensitivity analysis 2

The teachers incurred out of pocked (OOP) expenses, this was mainly to purchase books and materials to support the dissemination of formative assessment in their setting. The OOP were not mandatory but supported teachers to develop their knowledge and implement strategies. Thus, sensitivity analysis was conducted on the assumption that all teachers in the study incurred OOP expenses.

Sensitivity analysis 3:

The education provider (GwE) has a flat rate to cover supply cost for teachers. The flat rate is £250 per day and £125 for a half day. The cost of a teacher's time is not a standardised cost and could vary across different providers and may not represented an accurate cost. Thus, sensitivity analysis was conducted on the cost to cover teaching staff, using the costs calculated for teacher's time above in table 5.2.

<div align="center">

5.4 Results

</div>

In total eight teachers were interviewed, and seven focus groups were conducted with 57 pupils across the intervention arm. Table 5.3 details the participating pupils in the intervention and the control arm. The number of quantitative responses varies due to pupils being absent, missing data pre or post and pupils could choose to omit questions. In total 638 pupils took part in the surveys, 59% of the cases were omitted from the analysis due to incomplete post test data, as explained in Chapter 2 imputation is not recommended on one of the measures.

Table 5 3 Characteristics of pupils

	Intervention n=118	Control n=139
School year		
4	39	34
5	32	44
6	47	61
Total	118	139
Age		
8	35	28
9	31	37
10	42	47
11	10	27
Total	118	139

Data was analysed using SPSS 25 as the data was non normally distributed. However, given the size of the data set a parametric test was deemed appropriate (Blanca et al, 2017). A 2 x 2 mixed-model ANOVA was conducted on all measures. (See appendix D1 for the normality testing). Table 5.4 details the mean and standard deviation for all attainment measures. Table 5.5 details the results for the questionnaire data CHU-9D, SDQ and QoLS measures. No significant difference was found between all the outcome measures.

Table 5 4 Mean scores (SD), mean difference and p-values for, National test data 2018-2019.

Domains	Intervention group			Control group			Between groups	
	Baseline (*n* = 74)	Follow up	Mean change	Baseline (*n* = 77)	Follow up	Mean change	Mean difference	P-value
1. English age standardized score	104.36 (16.257)	103.85 (14.009)	-0.51	104.25 (11.768)	101.89 (11.171)	-2.36	-1.85	*p* = .206
2. English progress score	1006.11 (22.11)	1005.12 (21.846)	-0.99	1006.57(17.474)	1002.61 (16.247)	-3.96	-2.97	*p* =.133
3. Welsh age standardized score	100.81(15.310)	100.58 (15.468)	-0.23	103.78 (12.367)	105.25 (13.709)	1.47	1.24	*p* =.219
4. Welsh Progress score	1000.55 (20.812)	1001.77 (21.600)	1.22	1006.43(18.562)	1005.97 (18.860)	-0.46	0.76	*p* = .396
5. Math's age standardized score	106.30 (14.209)	104.47 (13.786)	-1.83	106.99 (15.897)	106.61 (14.195)	- 0.38	-1.45	*p* = .36
6. Math's Progress score	1009.41 (18.974)	1007.05 (17.770)	- 2.36	1009.61 (20.683)	1009.15 (18.197)	-0.46	-1.90	*p* =.340

Table 5 5 Mean scores (SD), mean difference and p-values for, CHU-9D, SDQ and QoLS questionnaires.

Domains	Intervention group			Control group			Between groups	
	Baseline	Follow up	Mean change	Baseline	Follow up	Mean change	Mean difference	P-value
1. CHU-9D	.885771 (.097594) (n=94)	.872000 (.1035466) (n=94)	-0.013771	.888340 (.089375)	.876812 (.0935050) (*n=110*	-0.0011528	-0.01261772	*p* =.868
2. SDQ	15.22 (3.956) (n=85)	15.04 (3.950) (n=85)	-0.18	15.23 (4.747) (n=92)	15.99 (4.249) (n=92)	0.76	0.94	*p* =.932
3. QoLS	3.48 (.326) (n=69)	3.36 (.449) (n=69)	-0.12	3.28 (.466) (n=69)	3.32(.381) (n=69)	0.04	-0.8	*p* =.135

Qualitative data.

Qualitative data was analysed using thematic analysis, for a full details of the procedure and coding see Chapter 2. Interviews were conducted with seven class teachers and the following themes were identified:

Teacher Interviews

Focused pupils own learning: All teachers detailed that the strategies they employed allowed the pupils to become more independent, they were able to not only get on task quicker, more ready to learn, but were able to self-assess and reflect on their work.

Understanding pupils learning: All teachers agreed that the strategies supported them to know where the pupils were in their learning and a better understanding of who the pupils are as people. Teachers were able to identify which pupils needed support much quicker, and also adapt teaching in the real time to support any misconceptions that pupils had within the learning process.

Self-efficacy: Teachers detailed that the pupils within their classes had developed more self-efficacy within the classroom, they were able to challenge themselves to do tasks they were hesitant at before. Teachers also found that pupils were more confident to contribute and share with the class.

Improved Behaviour: Every teacher that was interviewed commented on the dynamics of the classroom and how it had changed, the classroom was more inclusive and that pupils were not left out.

Reduced workloads: All teachers identified that the feedback was able to reduce the marking workload, mainly by improved standards of work, this in turn allowed them to mark quicker.

*Improved standards***:** The teachers all discussed that there was an improvement of standards, this fell in to two categories. Firstly, improved standards of work that was presented to them either with the use of success criteria, talk partners or peer and self-assessing. Secondly, teachers were able to see pupils have moved out of their predicted level, and this was particularly evident with underachieving pupils.

Prepared for the new curriculum: Teachers also saw that strategically formative assessment linked with, the 12 pedagogical principles and the four purposes set out in the Curriculum for Wales (Welsh Government, 2018), some felt that this project has prepared them for the role out of the new curriculum.

Focus Groups

Pupils were asked about their understanding of the strategies the class teacher had used over the academic year; they were also asked about the impacts they perceived. Full

analysis can be found in Chapter 4, the analysis identified the following themes: Focus groups were conducted with 57 pupils from eight different intervention schools.

Talk Partners: Pupils understood what a talk partner was, they were able to see that the strategy had other impacts, such as academic support and social relationships. Pupils discussed that one of the issues was copying, while some pupils did not identify it as copying, they identified that they would become annoyed at people looking at their work as well as the pairing being one sided with the workload being carried by only one of the pupils. Pupils were also able to articulate the frustration they sometimes feel with having talk partners, this came with certain pairings.

Mixed ability grouping and differentiation of task: Pupils were aware what 'group' they were in and if they were sitting with a higher or lower ability pupil and how this could support each other's learning. There were incidences where pupils explained that there were some negative consequences for both the higher and lower ability pupils. These strategies allowed them to be focused on their work and understand where they were within their own learning journey.

Elimination of comparative rewards: Pupils in two focus groups discussed how the rewards were being used as bribes for them to do work and monitor behaviour. Pupils were able to detail that the focus was not on their learning, but that the motivation was to get the reward. The pupils understood that without the reward system they would focus on the work they were doing and got feedback in other ways to support their learning and that removing the rewards made the classroom more equal.

Growth Mindset: Pupils in the focus groups were able to identify how having this strategy implemented in the classroom impacted on their work, self-efficacy, more positive and engaged with their work.

Success criteria: Pupils were able to describe how the success criteria was constructed, and that it helps them to complete a task successfully. Pupils discussed that the success criteria made it easier to stay on task, improve the standard of work and identify where improvements can be made. Mostly pupils discussed how the success criteria was co-constructed, a small number of pupils discussed the success criteria was not co-constructed.

Feedback: The feedback strategies allowed the pupils to see where they were within their own learning, identify mistakes and allowed them to be resources for each other.

Meta- cognition: Pupils discussed a shared language of learning, and how they were able to use this language in relation to tasks and how they could improve. Some pupils were able to use the shared language to retrospectively assess their work.

5.4.1 Observations:

The data on the observations suggest that in the intervention arm formative assessment principles were observed and are implemented either reasonably or well implemented see figure 5.5

Figure 5.1 1 Observational data
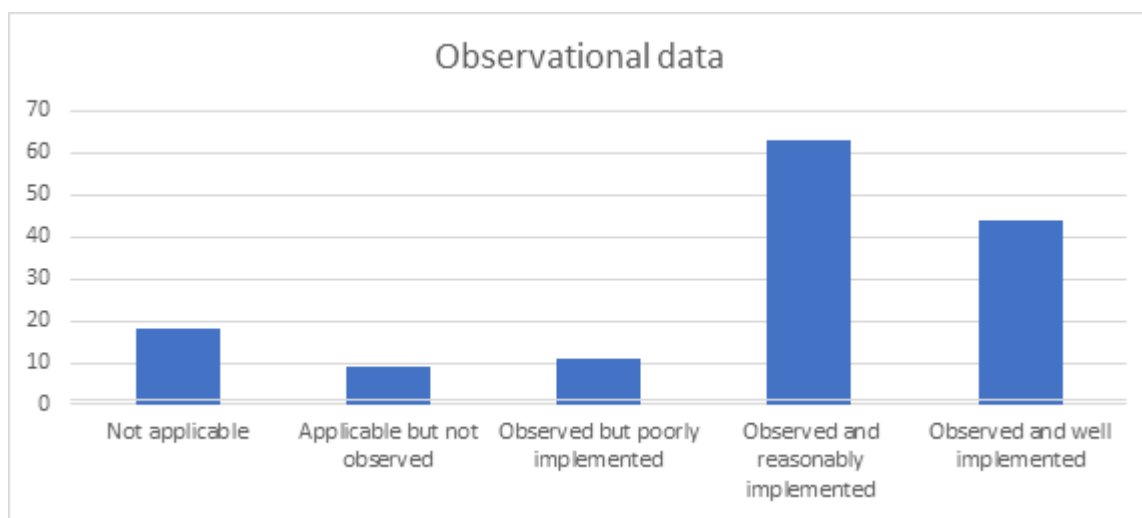
5.4.2 Costs:

Table 5.7 details the cost of the implementation of the North Wales Formative Assessment Project. Sensitivity analysis results appear at the bottom right of the table 5.7.

Table 5 6 Cost of FAIP

Teachers (Payment to school for release of the teachers for the training)				Inflated to 2022-23
		Units	Cost	
Training day 1	£ 250 per teacher	342	£85,500	
Training day 2	£ 250 per teacher	303.5	£75,875	
Review meetings 1	£125 per teacher	308	£38,500	
Review meeting 2	£125 per teacher	257	£32,125	
Showcase *		243	£22,356	
Staff costs				
Project manager (payments per day)	£750	70	£24,500	
Presenter and lead advisor (per day)	£350	25	£8,750	
6 Regional advisors (SC GwE team) 8 days	£350	8	£16,800	
5 Extra staff project members,	£350	1.5	£2,625	
Tier 1 teachers (lead and host review meetings)	£13,500	2	£27,000	
Tier 1 teachers for training days	£5,250	1	£5,250	
Expert Trainer	£3,000	1	£3,000	
General support of GwE Supporting Improvement Advisers with schools (1 day per school)	£350	193	£67,550	
Administration days	£103.13	50	£5,156.50	
Resources				
Venue (2 full days, 2 half days)	£38,189	1	£38,189	
Access to expert trainer Platform for GwE advisers	£250	1	£250	
Printing training materials	£1,611.73	1	£1,611.73	
Filming	£1,648.00	1	£1,648.00	
Translation (Materials and in person translation on training days)	£5,132.93	1	£5,132.93	

		Total	£456,919	£ 541,523
Teacher costs *				
Time (Time cancelled out by time saved)	£0.00			
Books	£355.00			
Materials	£0.00			
		Total	£355.00	
Intervention cost		Total	£457,274	
Number of pupils exposed to the intervention*	8075			
Cost per pupil			£56.63	£67.55
Sensitivity analysis 1 Class size Cost per pupil				
30	9690		£47	£55.70
20	6460		£71	£84.15
Sensitivity Analysis 2 Out of pocket expenses cost per pupil				
	£51		£58.63	£69.92
Sensitivity analysis 3 Buying out teacher's time cost per pupil				
£549,322/8075			£68.03	£80.59

* Showcase was attended by 243 teachers, even though GwE did not pay for the teachers there would have been a cost (opportunity cost for the teachers' time)
* Per pupil costs were calculated from the average teachers attending the training, two full training days were held 342teachers attended the first training and 304 attended the second with an average of 323
* Average of 323 teachers with an approx. class size of 25 = 8075 pupils

* Cost for teachers were collated from interviews with 7 teachers

5.5 Summary of findings

The cost-consequence results demonstrated that the total cost for the formative assessment project calculated to be £67.55 (converted to 2022-23 prices) per pupils based on the intervention delivered to tier two teachers only. This incorporated training 323 teachers. There was no statistically significant difference in the reading and numeracy attainment of pupils between the FAIP intervention schools and a group of matched control schools. There was also no statistically significant difference found on all other outcome measures used (CHU-9D, SDQ, QoSL). McMillan et al. (2013) discuss the difficulty to establish causality within formative assessment research, given the difficulties of the confounding variables that exist in researching in applied settings. The qualitative research undertaken with teachers and pupils in the FAIP schools found positive both perceived impacts in terms of teachers' skills and also pupil experiences from the teacher and the pupils. Teachers discussed mainly positive impacts on the classroom environment and improved standards of work. Some of the teachers discussed the reduction of their workload due to different feedback and marking systems. The pupils discussed learning to work in a different way, and they were able to focus on their work and understand what was expected from them in the tasks that were presented to them. While it is difficult to translate the qualitative findings into monetary values, the qualitative data suggests there were positive impact on provision and learning experiences

5.5.1 Comparisons with other literature

Attainment data from the EEF's 'Embedding Formative Assessment' study suggests that at GCSE pupils made two months progress as measured by their Attainment 8[20] scores (although there was not a significant difference in English and maths). Although the level of assessment is different, the results suggest that using measures of core subject progress, there was no difference following the introduction of a formative assessment intervention programme. The cost of the EEF 'Embedding Formative Assessment' implementation was £1.20 per pupil and was based on a two-year implementation averaged over three years. However, the EEF do not calculate the opportunity cost of teacher time and nor do they include detailed cost breakdowns, so drawing meaningful comparisons on costs is very difficult. Similarly, Yeh (2007) completed a cost-effectiveness analysis on 'Rapid Formative

---

[20] Attainment 8 scores are what pupils have achieved in their GCSE, whereas Progress 8 is the measures pupils against how they compared against others in pupils in England.

Assessment', and the cost to implement this brand of formative assessment was $28,31 (£28.44, converted and inflated to 2022-23 prices). However, this was discounted over seven years, and if we assume that the teachers retained the skills for seven years, a more meaningful comparative cost compared to FAIP would be £9.48 per pupil.

Research conducted by Ozan and Kincal (2018) identified similar outcomes to this evaluation, i.e., that there was no significant difference on their student achievement measure. However, their qualitative data similarly suggested that the implementation was positive., but research focused intensely on social studies classes, and could limit the findings to that setting. The difficulty with evaluating the impact of a formative assessment project is the mode in which training is received. The current FAIP utilised a specialist trainer and also a network of train-the-trainer support between schools, so comparisons are difficult due to design, population, and variation in the quality of implementation support available. Levin et, al (2017) suggest that this is a particular issue in educational research and when incorporating economic evaluations, comparisons have to be like-for-like.

The project STAR (class size reduction) is known as one of the biggest large scale state wide evaluations in education (Columbia University's Mailman School of Public Health, 2007), and has been of interest to policy makers and economists. Positive impacts were identified as a result of reducing class sizes from 25-20 to 13-17 learners (Reichardt, 2000). The, the cost of the intervention depended on the amount of students involved, for example, the cost of moving from a class of 24 to 20 learners was $435 per pupils (£636.43 converted and inflated to 2022-23 prices). While there is much discussion around the research design, effectiveness data and the way cost were collated, this large-scale intervention is a good example of how large-scale research for educational policy making can incorporate useful economic data (Justman, 2018). It is also a strategy that is clearly defined, using standardised outcome measures, and costed appropriately.

Ultimately, CCA provides the costs and the impact so that the decision maker can identify which impacts are most important to the context or setting (Hartfiel & Edwards, 2019). There is an underlying danger when policy makers only focus on educational effectiveness, reducing education to its most efficient point by overlooking the appropriateness of activities. While one intervention may increase standardised scores in academic domains, it should not replace other activities that have positive impact on other important outcomes such as learner well-being, enjoyment and participation in play as a mode of learning. Policymakers need to consider the wider curriculum and moral

implications of education provision when considering the implementation of new strategies or policies (Gilead, 2014).

5.5.2 Limitations

For the FAIP there is a relatively short evaluation window, universal interventions that are targeted at whole populations can take time for the population impacts to develop (Greenberg & Abenavoli, 2017) attainment impacts can be difficult to show over a short timescale. Kraft (2020) identifies that interventions that use measures designed specifically for the intervention will see large effects than interventions that use standardised tests (such as national test data). Yeh, (2007) '*Rapid formative Assessment'* used an intervention specific test and saw positive effect sizes. Perhaps a longer data collection period could have picked up on positive effects, or a formative assessment specific questionnaire could have been developed to pinpoint the impacts more effectively.

The sample size may not be representative of all the schools that took part in the whole interventions and this research only focuses on Years 4 to 6 in Key Stage 2. Teachers in the whole project came from primary, secondary, and special schools, and it would have been helpful to collect data from a wider range of teachers. However, given the timescale and limited resources this was not possible.

GwE school improvement officers carried out the structured observations of teachers delivering sessions to learners. While these officers are mostly ex teachers and have significant experience, it is possible they could have been biased towards wanting a more positive outcome for an intervention that their organisation was funding. Future research could include independent observers to reduce observer bias of using GwE staff. Also, it was not possible to assess inter-observer reliability in this study to evaluate the consistency between observers (Gisey et al., 2013).

One of the limitations is the estimation of the cost of teachers' time. Levin (2018) suggests that national prices should be used, as there are large disparities between rural and city school and pay scales differ between the home nations. There is no standardised cost for education activities, as such cost analysis could have been performed at national prices and not the cost that were borne by GwE.

Using a child health utility questionnaire has been used in school settings for various research, including weight loss interventions (Canaway et al., 2019) and more broadly in universal interventions tackling mental health issues in young people. However, the intervention under investigation here was not directly linked to mental or physical health. This is an important limitation. Wider interdisciplinary work needs to be carried out on the

Quality Adjusted Life Years (QAYLs) to be represented in wider areas so that robust economic evaluations like cost- utility analysis can be made (van IJzendoorn et al., 2020).

While qualitative data was able to demonstrate some positive impact for the perspective of the teachers and the pupils, using this information in an economic evaluation did little to support the economic rationale for FAIP. No impact was identified on standardised tests and there is a lack of shadow prices to transform the qualitative data into tangible 'prices' thus limiting the use of qualitative data in economic evaluations. With out this qualitative data, researchers risk reducing interventions to narrow outcomes.

5.5.3 Suggestions for future research

Further research over a longer timeframe than one academic year needs to be considered, to allow for formative assessment practices to embed into the classroom provision.

Given that the teachers were free to choose what they implemented in their class, further research could be more tightly designed around the strategies to give a focused view of the elements that were most successful in the classroom to inform future implementation.

More cost analysis needs to be completed in education in detail and include the opportunity cost so that decision makers have all the available information to make informed decisions. This includes the development of shadow prices to understand qualitative outcomes in a more quantitative way.

5.6 Conclusion

This research aimed to demonstrate how a cost analysis can be used in formative assessment and more broadly in education. Qualitative data showed that the intervention was received in a positive light by both the teachers and the pupils. While quantitative national test data and CHU-9D, SDQ and QoLS showed no significant difference. Longer term data could be tracked to identify trends longer term, and more focused tests may be suitable for the different elements. This research identifies a cost for teachers' time, and this is important for future costing of school interventions to represent more accurate costs and support costing for BAU in control conditions. Given that cost analysis has had limited use in education settings this research could support others to carry out cost analysis so that decision makers can make evidence and cost informed decisions.

Chapter 6

"It is essential then, that when dealing with educational productivity we take a broader perspective and keep in mind that education has its unique features that make not all forms of economic thinking adequate for it. We must at least consider the possibility that satisficing rather than maximising might be the right course of action when dealing with educational productivity and not uncritically rely on notions imported from mainstream economics." (Gilead, 2018, p. 280)

6.1 Introduction

This chapter aims to explore the methodological and theoretical considerations that researchers need to consider before undertaking cost analysis in educational research. While some guidance already exists (Levin et al., 2017), this chapter focuses on findings from the previous systematic review to identify where more research or considerations need to be made and understand the similarities and differences between health economics and the methods being developed in education. This comes at an opportune time where there are calls for more economic evaluations to be used in education (Kraft, 2020). Economic evaluation has the benefit of looking at the interactions of cost and effects to support decision makers but there are limited examples of this in education (Levin & Belfield, 2015). There needs to be a consideration of the landscape where the research is conducted.

While the rationale for this thesis has been to look to health economics, education, and public health face very similar issues. Public health and education both target wide sections of the population with associated challenges such as limited control over the dissemination. Interventions take time to embed (and the evaluation cycle is usually short Taylor-Robinson et al. (2008) and researchers must consider a range of outcomes (Weatherly et al., 2009), and this is also a difficulty faced in educational research (Ponce & Pagán-Maldonado, 2017). The analysis also needs to consider the complex economic and political landscape that decision makers are faced with (Ponce & Pagán-Maldonado, 2017). The global pandemic has had an unprecedented effect on the education systems, and this chapter aims to reinstate the argument that more economic evaluation is needed at a point where decision makers are faced with a practical problem (e.g., the covid learning gap) and additional support from governments (i.e., economic support) to support pupils. While education is moving towards more evidence-informed approaches, it is important to provide economic

evaluations, so the decision makers have all the information they require to help them make more balanced judgements (Levin & Belfield., 2015).

To get the best out of education provision we need to consider the efficiency in which resources are deployed through technical efficiency (maximising outcomes with the least amount of cost) and allocation efficiency (how the resources are used to get the outputs) Ruggiero (2007). Although this is no different to health where it is the focus of health economics to maximize efficiency, education like health does not function in the private market[21] where there are other drivers for efficiency. However, as researchers we need to understand the theory of economics; the landscape of the political forces that affect education policy; how and if increased spending in education has affected outcomes; and, the persistent issues in the education system. Finally, researchers need to understand how this translates to decisions being made within society and with public money.

6.2 Economic theory

Economic theory is seen in two strands, positive and normative. Positive economics looks objectively at events using price and quantity as economic variables known as the 'what is' and is value and judgement free. The normative strand deals with 'how it should be' in the economy, using an ethical framework to make decisions. Health economics comes under the normative strand and deals with what should be produced (supply and demand), how it should be produced (technical efficiency) and who should receive it (allocation efficiency), the demand for health care is infinite, and spending is restricted within a budget (Morris et al., 2012), thus decisions need to be made on supply, technical efficiency and allocation.

Health economists and policy makers work to maximise the efficiency and equity of distribution across society (Seixas, 2017). This is no different for education production where policy makers are concerned with technical efficiency, the best mix of inputs to get the best possible outputs and allocative efficiency maximising the outputs from a set of inputs (Gilead, 2018).

There are two main approaches with-in normative economics, welfarist and extra-welfarist. Decision about the desirability of different states of the world (individual preferences) are made through the welfarist lens where they seek to evaluate or construct a

---

[21] Apart from private health care and private education.

policy that will maximise societal utility[22]. Welfare economics wants to understand how goods and resources are delivered and the impact it has on the overall society. Welfare economics rejects the market model (selection, moral hazard, supplier demand issues) The framework or judgments in welfarist economics are based on the Pareto principle. The Pareto principle is a framework that seeks to understand an individual utility and aggregate this utility to social welfare thus to allow decision makers to allocate resources. If an individual is to gain 'utility' it is not at the detriment of another individuals 'utility' (Brouwer et al., 2008) under the Pareto principle.

The difficulty with using the Pareto principle is the judgments being made are concerned with social issues and there are a few issues with using the Pareto principle. As there are more than one individual in society, the difficulty is knowing what the individual's preference are. There is the assumption that people will want to maximise their utility by consuming goods or services, and there is no attempt to measure utility gained from helping others. Finally, the issue under the Pareto principle is that there is an assumption that everyone measures or values one unit or utility the same (Morris et al., 2012). The issue has become, not about improving one's utility but about making sure that utility is not forgone, and because of this difficulty and how to order/rank individuals' utility, creating a measurement issue. There has been no attempt to rank the Pareto efficiencies, so health economists have developed health utility related measures. This allows cardinal and individual preferences to be measured. This is done with health-related quality of life measure the most common one used in health is Quality Adjusted Life Years (QALY) a patient can attach preference values to their health status and used as a proxy to utility.

The second approach in normative economics is extra-welfarist, and moves beyond the Pareto principle, the sole focus is not on an individual's utility, and creates a wider evaluative space to incorporate other requirements. Under extra-welfarist economics the decision makers act as agents to the individual's utility, evaluating what will be best for the individual and is paternalistic in nature (Brouwer et al., 2008). Extra-welfarist can be utilised in other social policy sectors, poverty, housing education and so on (Seixas, 2017).

The extra-welfarist lens in the education system has been argued as the 'equity-efficiency trade off'. Policy makers will have to reflect the views of society and how much they are willing to trade off for the equity in their education system (Kislitsyn & Popova,

---

[22] Utility, broadly encompasses a person's preference to ordering of states or goods in the world, for a detailed discussion including how this term 'utility' is used in welfare economics see Brouwer et al. (2008)

2018). Understanding the theory of how economic decisions are made can lead us to see that the choice in running education systems are not strait forward and policy makers are tasked under the extra welfarist lenses of making education equitable for all. In general education has used cost benefit analysis as the monetary values to represent the utility and would be more associated with welfare economics rather than extra welfare economics as there are limited attempts in education to understand utility outside of monetary values (Gilead, 2018).

6.3 The education system

Public education like health must produce goods and services with limited resources, the demand for them is infinite, and thus the goods and services become scarce. Decisions need to be made in what to produce, how to produce the goods and services and how to share the goods and services to the population (Edwards & McIntosh, 2019). Economists in general deal with scarcity and choice, education economics thus uses the principles of economics in the education sector.

While education has changed since the industrial ages, education systems in the United Kingdom (UK) was the government's attempt to educate all children between 5-13 years old (Numata, 2003) to build their reading, writing and arithmetic Since the industrial era education became more secular (Gilead, 2018). Investment in human capital by governments through education was thought to yield returns to economic development, which is of importance to the stability of countries, (Lin, 1998).

Education traditionally has been concerned with the knowledge economy and come under two broad theories. The Human Capital theory, people who are educated have more valuable skills, are more productive and thus education has a causal relationship with the labour market. On a structural level a higher skilled workforce can increase gross domestic product (Pelinescu, 2015). Human capital investment can be seen to benefit the individual, through acquisition of skills leading to employment opportunities and higher wages. Another theory that is used to explain the private benefits to education is Signaling Theory (Rouse, 2017) that pupils who leave school with qualifications signals to the employer the person's ability (unobserved) are more able or motivated and obtaining qualifications. There are also spill over effects to society that run through both theories, in better health, less reliance on welfare and the intergenerational transfer of educational/ skills. This is important to government as they need to make judgments on where to invest resources so that aggregated returns can be sought from education over the long term for both the private individual (higher wages) and wider social benefits (GDP, health, welfare and so on).

The issue facing governments lies in the judgments they make on who and how people obtain education. The meritocratic approach is the most used method. This approach states that education "…underlies the prevalent belief that success in school and work is and ought to be determined by one's talents and efforts." (Mijs, 2016, p. 15) a person social background is irrelevant and fostering or discovering talent or ability requires equal access to education (equity). The investment in talent is seen in policies such as 'No child left behind act' giving every pupil the opportunity to have their talent realised (Mijs, 2016). The issue with a meritocratic system is the neoliberal economic policies that are surrounding the management of public spending including education. Neo-liberal policies can create inequalities in the system and works in opposition to the principles of the education system. While allowing for talent to be realised through equal opportunities, their still persists social inequality and a lack of social mobility linked to socio-economic status thus not truly meritocratic (Hall & Pulsford, 2019). Economic policy is not free from value judgments, and this is played out in the political policies that make education policy and spending (Viteritti, 2010).

6.3.1 The economic policy and the political influence.

The cost of education has been a primary concern since its modern universal approach, fraught with tensions from the political pendulum. A typified example of this is seen when education systems move to a neo liberal agenda, seen in the UK and the conservative government. While there were wide financial issues and recession, education spending came under criticism for not being able to solve societies 'ills', arguments around teacher accountability began to surface (Gillard, 2018). The concerns around education seen the system taking a 'consumer oriented' or 'quasi-market' model with the argument for more accountability in the system and more national testing, and this pointed the focus on individuals and away from systematic issues in society (Leyva, 2009). Neo-liberal policy in education saw the following changes, greater parental control with choice, competition through league tables, inspectorate framework, increased testing, rolling back the power of local authorities, government control over the curriculum and accountability of the teaching profession (Ball, 2016; Gilead, 2018). It is argued that this quasi-market model of the education system and the policy agenda has increased social inequality (Barone, 2019). In the United States of America (USA) push for more neo-liberal education policies came in the form of No Child Left Behind Act. The drive for efficiency in the education systems (along with a need for a skilled workforce to compete in the global market) had some disastrous

consequences; narrowing the curriculum (teach to test), greater inequalities were teaching was focused on pupils who would pass the test, and large amounts of students being categorised as needing remedial services thus not needing to take the tests (see Hursh, 2005 for a detailed analysis of neo-liberal agendas in USA education). Ball (2016) argues that there were other forms of neo-liberal policies such as external privatisation, through outsourcing of services to private companies, including professional development, school lunches, cleaning services and so on. This gave a real push to consider how educations systems are run, the efficiency and cost of supporting such education. There are many examples of neo liberal education policies, more recently, Charter Schools (USA), Swedish friskolor and Academies (UK) as a way of diversifying how public services are maintained and managed (Hall & Plusford, 2019), there seems to be a consensus among some research that relying on the market forces has resulted in little benefit or equity to learners (Eyles et al., 2017; Viteritti, 2010). This is a tension particularly felt in the English education system with the government wanting to academize more schools, despite the lack of evidence of effectiveness. Education systems have looked to the market model and while private education (fee paying schools) is able to command significant outcomes (Green, 2022), this is not the case for public schooling (Gorard, 2022). Health economics rejected the market model and provided a set of methodologies underpinned by economic theory that can distribute goods and services in the form of policy that seeks to improve equity, fairness and ultimately social welfare. The political agenda is an important context to how educations systems are run, but there are other important factors to consider, how spending has impacted outcomes, or where the spending is the most effective.

6.3.2 Spending in education

Other areas of societal spending are treated with great scrutiny, health, military and transport, economic evaluations have been present in these areas of spending since the 1960s to inform decisions on spending (Levin & Belfield, 2015) but this was not the same for education. There was growing concern about the cost of the education system failing, pupils leaving the education with little or no qualifications, and the wider cost to society and ultimately more public spending due to the increasing dropout rate. Levin and Bachman (1972) were commissioned by the American government to try to understand the wider societal cost of leaving school with no formal education. Levin and Bachman (1972) explored the cost in relation to increased crime, welfare reliance, health, and loss in tax revenue. This was one of the first reviews of its kind. In the work 'The Cost to the Nation of Inadequate

Education' Levin and Bachman (1972) set out to calculate how much money the government would need to invest to alleviate a growing problem in society as a consequence of a failing education system (the dropout rate). However, given this concern and a valid financial argument, many of the problems persist and the attainment gap consumes education systems around the globe (Schwab et al., 2021). Educational spending and additional funds made available for targeted interventions and groups of consumers have been a subject of debate, better qualified teachers, longer school days, additional tutoring and so on are all inputs that have been investigated to maximise educational outputs.

The Education Endowment Foundation analysis concluded that money did not affect attainment, the issue is being effective and efficient either financially or with their human resources (EEF, 2018) is what has an impact. To explore the issue of spending, researchers have made some advances, Return on Investment (ROI) evaluations propose the benefits of early education. Heckman (2012) is the expert in this field and argues that governments need to invest more into early education for the greater returns to pupils and society. Research on the High Scope Perry Preschool Program suggest the rate of return in 3:1 (Heckman et al., 2010) and programs like Head Start passing the cost-benefit test (Ludwig & Phillips, 2007). See Reynolds and Temple (2008) for an in-depth review of cost effectiveness or early years education. While early years education retains an evidence based economic argument, governments, policy makers, local authorities, school leaders are yet to have such informed answers for pupils across the system.

6.3.3 The difficulty of decision making in education

There are two broad issues that face decision makers, the first is trying to understand where to focus money and the second is to use research evidence to focus spending on interventions that will improve outcomes. This is to assume that there are good quality teachers in the classrooms, that are able to teach effectively but also engage with evidence and thus implement activities with fidelity. While not in the scope of this thesis, the need to recruit and retain good quality teachers should be at the forefront of any education system, then policy makers should grapple with the two issues of where and what to spend resources on.

To exemplify the difficulty for decision makers on where to focus spending, discussions will centre on targeted funding towards disadvantaged learners. It is a good example of educational policy that seeks to maximise outcomes with a particular population. As discussed in chapter 1 and 2 'the attainment gap' is a persistent issue all over the world

and is a policy focus for many governments (Mowat, 2018). In England, pupils from disadvantaged backgrounds are leaving school up to two years behind their peers (Andrews et al., 2017) and in the USA, the gap has changed very little over the last 50 years (Hanushek et al., 2020). The attainment gap has longstanding consequences for the individual, income disparities, poor health, entry into the criminal justice system and intergenerational transfer of poor attainment are just a few (Mowat, 2018). Within the UK socioeconomic status and in the US, ethnicity are the predictive elements in low attainment (Hadden, et al., 2020) and given the meritocracy principles billions of pounds in extra investment is targeted at disadvantaged pupils so that they have the chance to succeed to over com the income related poverty gap.

In the case of both the Pupil Development Grant (Wales) and the Pupil Premium Grant (England), schools were 'encouraged' to seek evidence-based interventions from organisations such as the EEF (Hilton, 2017) and school leaders have considerable autonomy to identify and employ interventions that suit the needs of their pupils (Crenna-Jennings, 2018). Carrier (2017) highlighted those interventions that are promoted to school leaders are disproportionally based on anecdotal rather than empirical evidence. Godard et al. (2020) highlighted that schools might struggle to elicit desired outcomes from effective strategies due to the challenges of ensuring fidelity of implementation (leading to ineffective practice). A review conducted by Pegram et al., (2022) highlighted that even when school leaders were made aware the interventions were not evidence based, they did not change provision. Education is facing a drive to become more evidence informed (Godard, 2020) the USA established the Institute of Education Science and the What Works Clearinghouse (WWC), and synchronously in the UK the EEF, through research, training and dissemination identify evidence to informed practice. So, while there is an understanding that increased spending does have an impact, in the case of funding that is targeted to disadvantaged learners the evidence is unclear (Pye, et al., 2015; Gorard, 2022). The inputs needed for disadvantaged learners, to acquire the outputs in line with others in the education system remain unclear and a difficulty for decision makers. So, to clarify the issues for decision makes is identifying 'what works'.

The second issue for decision makers, is research that is claiming to be cost effective (or low cost), but in reality, does not incorporate robust cost collection methods. Problems persist within research where claims of cost- effectiveness or 'virtually zero cost' is misleading, for example research carried out by (Vetter et al., 2020) proves positive results for the students but gives no indication to how much the intervention costs, there is no sense of how this intervention would be run without a research team and disseminated in a school

setting, materials, training and monitoring. While policy makers on the face of this sees a relatively simple way of raising attainment (as it is effective), it lacks the full picture of the consequences for them (financially). Hummel-Rossi and Ashdown (2002) argue that without identifying cost then evaluations are only partial evaluations, decision makers need to understand the interactions between the costs and outcomes from economic evaluations to support decisions to allocate funds (Isen, 2001). EEF detail that they only provide the additional expenditure or cost to implement and not incorporate the opportunity cost, *"This means that the Toolkit cost estimates do not reflect the total absolute cost of the different intervention types in the Toolkit, which would include the opportunity costs of using existing resources in a particular way."* (EEF,2018, p. 34). This could leave school or education providers implementing interventions that seem relatively cost effective but underestimate the staff time that would be involved to run the intervention. To clarify the second issue for decision makers is identifying 'how much'.

The issue of knowing what works at what cost continues to be the bridge that is still being built. How can educationalists or policy makers make informed decisions without a true representation of what is going on? How can they foster talent and ability if they don't have the evaluative tools? One way that this can be addressed is looking to the field of health economics and allocation of health care and the way in which they understand effectiveness and scarce resources Levin (2002). Health Economics have a wide range of methodologies to connect 'what works' to the 'how much' and there is growing movement for education to follow in the same path (Belfield & Levin, 2015; Levin, 2002).

The work of Henry Levin and his colleagues at The Centre for Benefit-Cost Studies of Education (CBCSE) have been championing the use of economic research methods in education to support decision makers (Bowden et al., 2020) although the uptake is slow and comes from a small base (Levin et al., 2018). Cost Effectiveness Analysis (CEA) and Cost Benefit Analysis (CBA) are gaining popularity to support decision makers on all levels of the system (Levin & Benfield, 2015). Godard (2020) suggests that without exploring what works in education, money will continue to be wasted in ineffective areas. While this move can seek to ask, '*what works'* policy makers are still left with the *'at what cost'* question and where best to invest resources for maximising outcomes.

6.4 Why now

Although the argument for economic evaluation in education is not new, there are new challenges facing the education system at all levels. While the globe recovers from a

health crisis 1.6 billion pupils were not in school due to enforced lockdowns, policy makers and governments had to adapt to the changing provision (Kaffenberger, 2021). The use of emergency education (Kagawa, 2005) meant education being delivered remotely. This ultimately had an impact on attainment e.g., 3 months learning loss in Key Stage 2 and up to 7 months learning loss for pupils from disadvantaged backgrounds (Twist et al., 2022). Research by Fuchs-Schündeln et al. (2022) suggest reduced investment in human capital (over the long term) because of COVID-19 will have long term effects on pupils' futures and suggest that even wider effects on non-cognitive outcomes.

The response from the education system was to provide additional funding to support pupils to 'catch up'. Perhaps now is the time that we realise that the lack of economic evaluations makes it difficult for governments and education ministers to identify cost effective strategies to support large-scale catch-up plans. Figure 6.1 details the additional funding allocated to schools for COVID catch up (Sibieta, 2021). The UK government catch up plan for pupils in England came with limited guidance, school leaders were asked to use their local knowledge of the area, the pupils and encouraged school leaders to seek information from the EEF (Sibieta, 2021). Thus, allowing the decisions to be left to the school leaders (Harmey & Moss, 2020), however research suggests that that spending in schools has not been effective and some interventions are not evidence based (Kraft, 2020; Pegram et al., 2022).

Zhou et al. (2021) discuss the learning loss and associated wider issues would need an additional $12,000 per pupil (£8,490) over five years. Analysis by Kraft and Falken (2020) propose catch up plans using extended days and tutoring has some evidence base however the cost analysis is not clear, the effectiveness data they use for the tutoring program from Ander et al., (2016) does not have a full economic evaluation or detailed costs, so could be misleading. Kaffenberger (2021) demonstrated an insightful modelling of rebuilding education systems post COVID, but there is no mention in the paper of the cost. Given the history of concerns of efficiency and outputs in education, perhaps now is the time to re-evaluate economic evaluations in education that will support decision makers.

While it is imperative that the government support pupils and provide additional funding to schools, what do these arbitrary values actually mean? How can we support school leaders, policy makers and government officials to target the funding where it is needed? The two issues are still being faced as mentioned above where to focus the additional funds and what is the best available evidence that incorporates a full evaluation (including a cost analysis).

Figure 6. 1 Spending amount in UK for catch up in education. (Sibieta, 2021)
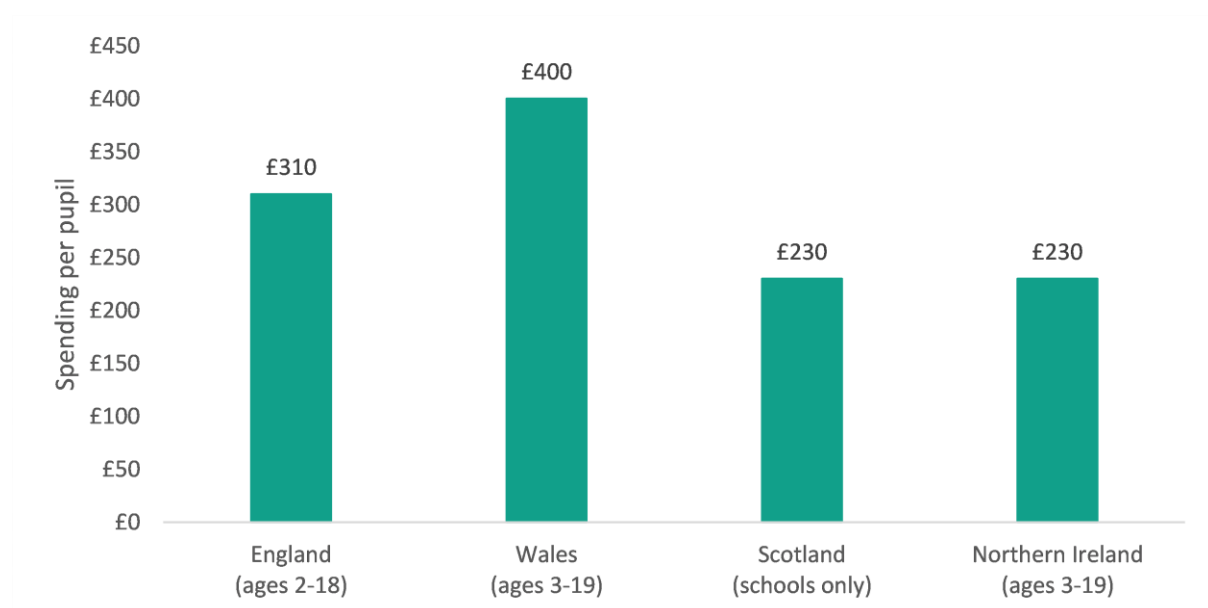
Looking to the health discipline who have similar pressures from society, allocating goods and services to mass sections of the population within limited budgets. Economic evaluations are more prevalent along with decision making frameworks health economics makes a good base for education to understand what is involved (Levin, et al., 2017). The second half of this chapter will map what an analyst practically needs to consider before undertaking a cost analysis.

6.4 Part two

6.4.1 Informing decision makers:

Fundamentally the reason to conduct any cost analysis is to support decision makers to make informed decisions with correct data so that they can maximise the outcomes or have efficiency in the system. Levin and Belfield (2015) suggest there are two criteria that need to be observed so that informed decisions can be made, first that the evaluation method used is credible and cost analysis is robust. Secondly that the information presented to the decision makers is understandable, and useful. A cost effectiveness ratio (CER) shows the incremental

effect the intervention has over business as usual (BAU) and is presented as per pound spent to gain the incremental effect. To take this further the marginal effects can be calculated by using Incremental Cost Effectiveness Ratio (ICER), this could help decision makers to decide between treatments. Chapter 3 details that the presentation of outcomes particularly when reporting CEA studies in education there was inconsistencies, some reported cost per unit of effect size (like traditionally done in health) or per pound effect size. A study by Yeh, (2007) reports relative cost-effective ratio this is not a recognised reporting of economic evaluations and only serves to confuse a relatively underdeveloped method. Other researchers Bowden et al (2015) calculated an effectiveness cost ratio and again this is not a common reporting method, having common reporting methods would solidify comparisons and make it easier to disseminate to decision makers (Barret et al., 2020). Clarity in the reporting would support understanding, and be digestible, Levin et al. (2018) suggest that the presentations of ratios should be consistent like in health. The analyst needs to be aware of the audience, is the analysis going to be understandable, so that it is meaningful, and the findings are purposeful. This needs to also be considered at the level of technical knowledge of the decision maker. While policy makers may be able to understand high level technical economic language, school leaders and policy makers may not be so accustomed to technical language (Lundkvist et al., 2021).

6.4.2 Opportunity costs:

Opportunity cost is central to economics it is different to accounting where the costs are explicit and are collected when goods and services are exchanged (Morris et al., 2012). Opportunity cost is the value of benefits foregone by not using resources in their next best alternative use and is seen as one of the most central concepts in cost analysis. While including opportunity cost can be relative to the perspective, some researchers in health argue that opportunity cost should be included regardless of the perspective as it can support decision makers (Lomas et al., 2022; Levin et al., 2017) and argue that in education opportunity cost should be included so that decision makers understand the true cost to replicate activities.

One of the main opportunity costs in education would be the opportunity cost of the teacher's time, an example would be, a teacher is trained in a different mathematics intervention, they are forgoing the alternative maths intervention and thus should incorporate the opportunity cost of this time. While excluding the opportunity cost could be because of

the perspective the analyst takes, it is important to consider when costing activities in education. There are examples of educational research attempt to collate costs or claim to be cost-effective as mentioned in the first half of the chapter. Primarily they will detail the cost of training and materials, but neglect teacher time, and other personnel needed to support the intervention. Machin and McNally (2008) present the cost of The National Literacy hour, but the costs are not described in any detail, just what the programme was planned to cost, although they do suggest that the cost may not be accurate and a true reflection, they do proceed to calculate a 'back of envelope' analysis (Kraft,2020, p. 236) cost benefit analysis and suggest that the intervention is cost-effective. Some authors make claims, "Since physically-active lessons can be teacher-led and use minimal resources, they are likely widely translatable and cost effective." (Vetter et al, 2020, p 735) and yet do not provide any cost analysis to support cost-effectiveness. Similarly, Lynch, et al. (2020) claim cost-effectiveness but present no cost data, it is misleading to propose cost-effectiveness without the relevant methods to collect the data or the methodologies to calculate ratios/ ICERS. While it is not to question the effectiveness of studies, questioning how costs were collated and importantly how they then can support financial decision making with scarce resources needs to be clear and transparent. Government endorsed organisations like the EEF provide a costing tool kit that presents costings for programmes but aforementioned do not incorporate the opportunity cost, and although they present empirical findings in a digestible way, there are some criticisms around how they use evidence (Wrigley, 2018). For a decision maker to implement interventions they will need to consider the staff costs (opportunity cost) to understand the true cost of the intervention (Institute of Education Sciences. 2020). Essentially nothing is free, even with the classroom already in a school with teachers, calculation costing needs to be done as if you were going to start from the beginning (Institute of Education Sciences, 2020). While the analysist my take on a narrow perspective to collate costs, opportunity cost particularly staff costs need to be made clear if they are included or not.

6.4.4 Cost collecting methodology:

As discussed in Chapter 2 there two ways to collect cost, bottom up or top down. Different ways of costing produce very different results in the cost outcomes, it is the job of the analyst to determine the correct cost collecting methodology, for example if the analysis is not concerned with variation of local cost, then top down would be suited, but if the variation on a local level is a consideration, then bottom up may be suited (Chapko et al., 2009). Within the education literature the cost is not centred on the unit of output, and within

health economics the cost collecting methodology centres on the perspective. Levin et al. (2017) suggests that for evaluations in education the analysis must include all the costs that are associated with delivering the intervention, this would be more set within the bottom-up costing.

6.4.5 Perspective.

Another element to consider when collecting costs is the perspective, which position the analysis takes will depend on what costs will be collected, for example in public health it is recommended a public sector perspective is used, for the technology appraisals the NHS and personal social services perspective is used (Jones et al., 2019). If the perspective is from the employed/patient perspective and is focused more on the cost to the individual. Table 6.1 gives examples of the different costs in a clinical and non-clinical setting that may be considered.

Table 6. 1 Differences in types of cost in a clinical and non-clinical setting, (Jones et al., 2019).

Type of cost	Description	Clinical setting example (GP clinic)	Non-clinical setting example (school)
Direct	Costs that are directly associated with the programme under evaluation (e.g. equipment)	Staff costs (both monetary and opportunity costs) for a nurse to deliver a smoking cessation service	Teacher time (both monetary and opportunity costs) to deliver an anti-bullying intervention
Indirect	Costs that are associated with the programme under evaluation, but are not directly attributable	Loss of earnings for the individual to attend the smoking cessation service appointment during working hours	Parental loss of wages to meet with school to discuss bullying incidents
Intangible	Costs that are associated with concepts that are difficult to quantify and measure (e.g. pain and anxiety)	Cravings during first few days of nicotine withdrawal	Emotional distress of child who is being bullied
Marginal	Costs of providing one more unit of a good or service	Treating one additional person at the smoking cessation service	Rolling out the anti-bullying intervention to an additional year group in the school
Incremental	Additional costs incurred by one service compared to another	The additional cost of providing the smoking cessation service by a nurse compared to providing the service through a smoking cessation leaflet service	The additional cost of providing the anti-bullying intervention by a school teacher compared to providing the intervention through a mobile phone app or other online service

Within education Levin et al. (2017) has developed 'The Ingredients Method' which is close in nature to the 'bottom up' approach in health economics. Costs are calculated on how the project was run and all the elements that are needed to run the intervention including the opportunity cost. Levin et al. (2017) suggests that cost should be calculated at national prices, not local this is where the cost methodology differs with health economics. Levin and Belfield (2015) suggest that were possible costs should be collected over various sites to accurately calculate the cost, however that is not always possible or viable to run multi-site experiments. The guidance in education in collecting cost are not split into different types of costs like health economics, direct, indirect, and intangible cost are collected where possible, but cost is calculated at national prices.

6.4.6 Business as Usual (BAU).

Interventions are usually compared to 'business as usual' (BAU) and cost need to be collected for BAU to use as a comparator to the intervention. For effectiveness studies this could be the cost of the control group. Health researchers can utilise the resource Personal Social Services Research Unit (PSSRU) which is a detailed unit cost resource for researchers when they are costing up interventions and treatments (Curtis, 2013). PSSRU, is similar to a cost index, the analyst for example would want to find out how much a doctors consultation visit would be (including the oncosts not just the salaried cost per hour), the PSSRU details this. For education it is not as straightforward Scammacca et al. (2020) and Hollands et al. (2016) have faced this issue when trying to cost BAU. When an intervention is not an added activity like after school, and it is a change to a teaching provision with a different pedagogical style collating the costs for BAU and the intervention costs could be difficult. Chapter 5 details the difficulty in collating the cost of BAU even for the basic cost of having a teacher in a classroom, while it might be practical to understand how much a teacher costs based on average wages it does not incorporate the cost of the teacher to the system, including pension and national insurance contributions. A teacher costs more to the system for example: on costs of having a clean and maintained classroom (Shand & Bowden, 2021) pension contributions and so on. The Centre for Benefit-Cost Studies of Education have developed a costing tool to quantify costs to support researchers and educationalists costing interventions although this is only in USA the development of this is a step froward. Shand and Bowden (2021) have attempted to standardise some of the assumptions that an analyst

has to make. Although they are from America, they are relevant to move the method forward and create a centralised repository for researchers. When considering changes to teaching practices that are incorporated into normal provision (training might be needed), both control and interventions will need to be costed in. Having a centralised cost repository like the PSSRU would support the analyst.

6.4.7 Time horizon

Activities in education need to consider the time frame for analysis, while there are CBAs detailing the higher wages and impacts over the life course, analysts need to consider how long the impacts will take to come to fruition. Given that education and public health share some of the same issues when considering impacts. Small studies can have positive impacts (Kraft, 2020) but as they are scaled up the impacts are reduced, and this is also evident in public health (Taylor-Robinson, 2008). Policy makers need to access evaluations that are over a longer time where the impacts can be realised in a difficult research landscape. Without the longer scale evaluations policy makers risk making decisions based on small scale intervention that will not have the same impact over a longer period or when scaled up (Buxton, 2011). Behaviour changes interventions (common in public health) take time to come to fruition (Michie et al., 2018) as with in education, impacts may not be fully realised for effective decision making (Murray, 2019).

6.5 Considerations for education economic evaluations.

6.5.1 Cost collected at time of intervention.

Difficulties can accrue when collating cost retrospectively, cost evaluation should be done at the time the intervention takes place. This is common in the more developed area of health economics where costings are built into pilots, trials, and effectiveness studies to whole populations (particularly public health [Vassall et al., 2017]) and is recommended in educational settings (Levin et al., 2017).

6.5.2 Effectiveness measures and populations must be similar or have common goals.

Measures: For use in cost-effectiveness outcome measures must be comparable or have similar goals, without similar outcomes this could be difficult to decision makers. Research that compares reading outcomes and mathematic outcomes would not be comparable and provide limited information to policy makers.

Population: Populations need to be comparable, if effectiveness is sought on equivalent demographics more reliable conclusions can be made. For example, comparing and intervention where the cohort is in an area of high deprivation to a cohort in a more affluent area. This will limit the conclusions being drawn.

In education standardised test have been utilised however they may not measure all the benefits of an intervention and could risk narrowing the curriculum. Even in the case of CBA attention needs to be directed to the population and outcomes. Clear and detailed reporting will support decision makers to use the resources effectively without missing relevant information. Analyst must consider if one intervention is comparable to another. One way that Levin and Belfield (2015) have proposed to mitigate the issues is to involve the stakeholders in the evaluative framework, this has also been evident in public health where there are difficulties evaluating interventions that target large parts of the population and can encompass various impacts (Fotaki, 2010).

6.5.3 Collect data from multi-sites.

Interventions do not react the same way in different places, so collecting the effectiveness and cost from multi-sites will give greater accuracy to the data. Hollands et al. (2016) also suggest that multi-site data can help interrogate fidelity. This is prevalent when considering interventions run in rural areas, where additional resources may be needed for example, travel consideration, recruiting staff. While in more populated areas such as cities hiring buildings, and equipment may incur higher costs.

6.5.4 Costs should be reflective of national prices.

Collating costs of interventions at local levels gives some accuracy to the cost of an intervention it does not incorporate wider market prices, interventions run in rural or urban area could produce very different costs. National prices could provide a clearer picture to the over all cost and is transferable to other areas.

6.5.5 Apply sensitivity analysis

Uncertainty in costs, effects the assumptions the analyst makes can be somewhat mitigated by conducting sensitivity analysis. Within health research this has been a

longstanding component of the analysis. Given that costs can be variable researchers need to be transparent about how they have identified costs and what assumptions have been made (Limwattananon, 2014).

6.5.6 Widen the evaluation field

The health discipline is experienced at conducting economic evaluations, exploring working in a cross-discipline manor could move the method forward. Education can optimise on the work that has already been carried out in health cross disciplinary research and method borrowing has been seen as adventitious. As with other disciplines there is the consensus that through rigorous research designs, methods and even theory borrowing can seek to understand social phenomena (Fellows et al., 2020).

6.5.7 Training

More training is needed to bridge the knowledge gap, training more researchers, policy makers and the education workforce will support the development and understanding of economic evaluations. Cost collecting tools like that from the Centre for Cost Benefit Studies in Education, or the Education Endowment Foundation need to be made available to a wider population. Funding within the system needs to address the gap so that rigorous training can be deployed and utilised. This needs to be for researchers to build up cost analysis but also for decision makers to utilise findings (Levin & Belfield, 2015).

6.5.8 Consider the journey to evidence:

While there is a shift to use evidence informed practice, within the research community there is also a drive to support this movement with the journey to evidence (Owen et al., 2022). As research develops through the stages, from inception of a research idea or the basic science, through to efficacy, effectiveness all the way to maintenance studies, cost should be considered through all the stages of getting evidence into practice. This is crucial to the development of treatments in the health sector where cost evaluations are built into the evaluation framework at all stages of the trials.

6.6 Conclusion:

Levin et al. (2017) book '*Economic Evaluation in Education'* gives clear guidance and rationale to the inexperienced researcher/ educationalist a way to approach economic evaluation. The EEF have an evaluation guide for teachers and school leaders to use but do not discuss the costs of the intervention or how a teacher would calculate such analysis (Major & Higgins, 2019). While this might be a above on beyond what is expected for a teacher to evaluate in their practice, not considering any cost again leaves the evaluative framework short (Hummel-Rossi & Ashdown, 2002) The Centre for Benefit-Cost Studies of Education have developed a *CostOut* tool to support researchers and educationalists to do economic evaluations (Hollands et al., 2015), this can support schools in conducting evaluations using economic methods and support translating the results into practical and usable results (Barrett et al., 2020). However further suggestions need consideration to build a robust framework to continue supporting both high level decision makers and close to practice decision makers.

While this chapter wanted to reframe the argument that to get the best out of education, we need to consider the efficiency, either through technical efficiency, allocation efficiency and to do this cost analysis is needed to support decision makers (Hollands et al., 2016; Levin & Bachman, 1972). The recovery from COVID-19 in the education sector helps to bring up this important issue; to support decision makers they need viable options, otherwise as researchers we are letting them down and the pupils they serve. The old tradition of 'throwing money at the problem' (Dickson & Harmon, 2011) does not seemed to have work for disadvantages learners, so we really need to consider more sophisticated ways to support decision makers.

Policy makers and governments need to also consider funding more training for educationalists to learn about the methodology or economists need to support education and apply the skills they have while understanding the idiosyncrasies of educational research (Levin, 2002), whichever way round more needs to be done to support our education systems so that there is efficiency, equality and transparency in decision making.

While there are differences between the different disciplines of economic evaluations the theory that underpins them is broadly similar. While it is still relatively early in the journey, as education takes a stronger focus on economic evaluations. There is the opportunity to develop the methods, work out the differences and the direction that education research wants to pursue (Gilead, 2018). While the bridge is still under construction between the two methodologies, this chapter serves to strengthen the links and highlight the

differences in the methodologies, so that there can be movement forward and support pupils reaching their true potential.

The development and drive for education to be evidence informed, with economic evaluations to support decision makers is well underway. There also needs to be a consideration to some of the theories that support this type of analysis, while not yet fully realised, the difficulty for education is to translate, define and set preferences around 'utility. The difficulty around 'utility' may lead to CBA being used more as it could be easier and more palatable for decision makers to understand 'utility' in monetary values (Gilead, 2012, 2018). The future work outside of the methodological considerations is perhaps to dig deeper and understand what 'utility' is in education and how this would be measured. To understand the theories that underpin each discipline will play out in how the methods and research design with be disseminated (Fellows et al., 2020). Developing a clearer definition of what 'utility' is must come with caution, as within health it can lead to a narrow set of outcomes and that could similarly focus outcomes in education to narrow a view of what education should do and not do.

Chapter 7

"For researchers, determining the cost-effectiveness of the programs they are developing is a reality check on the scalability of an experimental program that has demonstrated efficacy. Therefore, making CEAs a standard part of efficacy studies is in the best interest of both the research and practice communities." (Scammacca et al., 2020, p.384)

7.1 Discussion

7.1.1 Purpose of the thesis restated

Over recent years there has been a drive to create a more evidence-informed education system to improve the quality of provision and outcomes for learners. However, to better inform decision makers, considerations around the costs of different courses of action need to be incorporated into evaluations (Hummel-Rossi & Ashdown, 2002; Kraft, 2020; Levin & Benfield, 2015). Identifying value for money in educational spending can support policy makers effectively use public money (OECD, 2022) in the addition to good quality teachers that are effective in raising the outcomes for pupils (Hattie, 2003; EEF, 2021). Including cost analysis has been discussed as the next step in the journey for education systems to be evidence informed (Detrich, 2020; Owen et al., 2022).

7.1.2 Findings by chapter

Chapter 2 described the study design for the FAIP intervention. Given the scale of the project, the empirical studies were built into a Theory of Change framework (Chen, 2006). This chapter set out the rationale behind the data collection methods and provided the justification for the empirical studies contained in the thesis.

Chapter 3 was a systematic review to establish the use of methods of health economics in education. A total of 12 published full economic reviews were identified; eight of these were cost-effectiveness studies and the remaining four were cost-benefit studies. Measures ranged from standardised tests, high school completion and researcher designed tests. Using the Drummond et al. (2015) checklist to examine the quality of the included studies, three were judged to be poor quality, five were judged to be moderate quality and the remaining four were judged to be good quality. Technical issues including the choice of comparators, cost analysis, and the presentation of findings limit the degree to which these studies can be judged helpful in supporting decision making processes in education. There was evidence that economic methods have been used to evaluate activities in a school setting,

the most common method was cost-effectiveness analysis, followed by four cost benefit studies. The following recommendations were identified from the systematic review.

1. Before commencing efficacy or effectiveness trials, researchers should consider costing an intervention and/or programme at an early stage.

2. In the case CEA evaluations, the choice of comparator is an important factor to ensure that meaningful conclusions can be drawn between studies. Outcome indicators and scales should be considered when comparing interventions.

3. Cost-effectiveness ratios should always be included to ensure uniformity in the presentation of information in the discipline of education economics, thus making it easier for decision makers to interpret study findings.

4. Where possible, a common indicator of effectiveness would enhance the comparability of interventions.

5.While this review used a critical appraisal tool developed for health economics research, education research that uses economic analysis would benefit from the development of a bespoke tool to strengthen the quality and standards needed for education economic evaluations.

Chapter 4 built on the evaluation framework that was needed to understand the impact that the FAIP had from the teachers and pupils' perception. The chapter contained two empirical studies: (1) interviews with class teachers (Study 2); and (2) focus group interviews with pupils in the intervention group (Study 3).

Study 2 aimed to answer the following sub-research questions:

1. What are the perception and experiences of Tier 2 teachers in relation to implementing formative assessment strategies and the perceived impacts in the classroom?

2. What are the barriers and facilitators of implementing formative assessment in the classroom?

A total of seven teachers were interviewed for the second tier of the FAIP, and the results indicate a range of positive impacts within the classroom. Some of the benefits included improved behaviour, including the fact that teachers felt they better understood where the pupils were within their learning and understood them more fully as individual learners. Some of the teachers discussed that the pupils were returning work of a better

standard and that the feedback was reducing their workload. The pupils were more focused on the tasks and that they had better self-efficacy.

In terms of implementing the strategies the teachers discussed some barriers that needed to be addressed. There was some resistance; for the pupils this mainly centred on not wishing to sit with peers outside of their friendship group or with a different gender. Some teachers faced resistance from parents particularly where the strategies were not being implemented in all the classes across the school. There was also some evidence to suggest resistance from other staff members due to having a different pedagogical ethos in their classrooms. The teachers discussed the facilitators of implementation: a supportive SLT allowing them the freedom to disseminate the strategies; and presenting to the pupils and parents so they had an understanding as to what the strategies and the rationale behind the changes were. The teachers also discussed adapting the strategies to fit their context. Overall, the majority of the teachers' perceptions were positive, and they were able to disseminate the strategies. Future research needs to focus on some considerations around how the barriers might be mitigated, for example, supporting teachers to communicate the strategies to pupils, parents, and other staff in school.

Study 3 aimed to answer the following sup research question:

1. What are pupils' experiences and perceptions of using a range of formative assessment strategies?

In total 57 pupils from eight schools participated in the focus groups. Pupils were chosen at random by the class teacher, and all had parental consent to take part. The pupils ages ranged from 8 to11 years (in Year 4, 5 or 6). The pupils were able to understand the strategies that the teacher had disseminated, the pupils could articulate the strategies were all to focus on the work that they were producing. Some pupils were able to discuss the benefits of working with other pupils outside of their peer groups. The pupils were able to discuss what the outcomes of the strategies were and showed a good level of understanding.

There were some pupils who discussed that some of the strategies, particularly talk partners, presented them with some challenges. Also, the random seating arrangements meant pupils were sat with other pupils outside of their peer groups. Clarke (2014) suggest that teachers need to create a positive classroom environment to avoid issues with talk partners. One issue that was raised was the concept of pupils copying each other's work, and pupils being aware of the ability what ability group they were in, and for the lower ability pupils having negative feeling towards themselves. This element needs to be explored further to support the implementation of talk partners.

Chapter 5: An evaluation of the impact of the FAIP intervention showed no statistically significant differences on all pre-post scores between the intervention and control groups. The cost consequence analysis (CCA) of the FAIP for Tier 2 teachers was £541, 523 and this equated to £67.55 per-pupil (inflated to 2022-23 prices) in the intervention condition. However, incorporating the qualitative findings demonstrated a positive impact from the teachers and pupils' perspectives. The pupils were able to understand and articulate the strategies that were deployed in the classroom. From the observational data, the majority of the strategies were reasonably or well implemented. The rationale for using CCA was that it gives the decision maker all the information, so that they can identify the elements that are most important to their context (Glouberman & Zimmerman 2002; Rodgers et al., 2008).

Chapter 6: This chapter detailed the theory behind health economics and the ethical framework that this discipline sits within. The chapter discussed the difference in health economics and the economic framework being developed in education. The chapter draws on some of the main findings identified from the systematic review.

This chapter highlighted the core elements that a researcher needs to consider when taking on and economic evaluation. The purpose of an evaluation should be to support decision makers, so the researcher needs to articulate the way the economic evaluation was conducted. The researcher needs to present the findings so that decision makers can interpret the results in a context specific and meaningful way. The opportunity cost should be routinely considered in evaluations, and this contrasts with many existing economic evaluations used in education settings. Within health economics the opportunity cost would only be included if it was relevant to the perspective of the analyst undertaking the evaluation. Given the recent developments in the education discipline around economic evaluations, opportunity cost should always be included. The need to calculate bottom-up costings, including all the cost associated with running an intervention and to collect costs at the national prices is where there is a diversion in the different disciplines. There is also a need to develop more robust business as usual costs to develop the cost collecting methodologies. Table 7.1 details the considerations for the research before embarking on a cost analysis.

Table 7. 1 Consideration for the take up economic evaluations.

Cost collected at time of intervention	Where possible collect the cost when the intervention is happening.

Have common goals.	Outcomes and populations need to be the same or very similar to make comparisons.
Collect data from multi-sites	Where possible collect cost from different sites to identify implementation differences.
Costs should be reflective of national prices.	Collecting national prices might give a more robust cost to implement interventions in other areas. Local prices can vary substantially.
Apply sensitivity analysis	Make clear the assumptions the analyst has made and test these assumptions.
Widen the evaluation field	Develop cross discipline working; education can optimise on the work already developed in health.
Training	Training for researchers, to carry out more evaluations and training for decision makers to interpret findings.
Consider the journey to evidence:	Build evaluations into the journey to evidence (discussed in detail below).

7.2 Research questions revisited

RQ1: What is the evidence for the effective use of health economics approaches to evaluate education programmes/provision?

There were limited examples of health economics approaches being used in education. However, there were some good quality examples being carried out Bowden and Belfield (2015) Hollands et al. (2016) Belfield et al. (2015) all conducted high quality cost-effectiveness analysis and Reynolds et al. (2002) conducted a high-quality cost benefit analysis. One of the main limitations identified was the lack of consistency in the outcome indicators used in across evaluation studies, thus creating difficulties comparing different interventions. There were also differences in the ratio calculations particularly the incremental cost-effectiveness ratios leading to limitations on the conclusions that can be drawn for the analysis. Chapter 3 did identify there is already work underway in education around economic evaluations that is not seated in health economics. Particularly the 'Ingredients method' developed by Levin (1988). The work being carried out in America in The Centre for Benefit Cost-Studies in Education featured heavily in the systematic review and demonstrates that the field of education is developing its own methods around economic evaluation. Chapter 6 discusses some of the main differences between health economics and

the new field of economic evaluations in education and the considerations needed around using the methods.

 With reference to research question one, there are some examples of health economics being used in education but there are more examples of economic evaluations being used in education that are built on new and specific methods being developed in the field of education. Critically, while most of the included studies used the 'ingredients method' to collate costs, eight studies were poor to moderate quality suggesting the appraisal tool may not have been the best fit for the included studies or that there needs to be an impetus for better quality economic evaluation in education, regardless of the cost collecting methodology.

RQ2: What is the impact of a regional formative assessment programme, and does this represent value for money?

 An evaluation of the impact of the FAIP intervention showed no statistically significant differences on all pre-post measures between the intervention and control groups. However, the qualitative data from the teachers and pupils does demonstrate that there was a positive impact following the delivery of FAIP. The participants that took part in this research saw this as positive initiative that helped their learning. In the following summary, the main findings from the qualitative research are organised around the five central principles of effective formative assessment:

 1. **Sharing Learning Expectations: What the pupils are going to learn and the success criteria to achieve the learning**: The use of the formative assessment strategies particularly the success criteria were an effective tool in helping the pupils to understand what was expected from the task and if the task was challenging them. Pupils were able to detail that the success criteria made the task they were completing more strait forward. They were able to check and improve the task that they were completing.

 2. **Questioning: Effective questioning to facilitate learning:** All teachers agreed that the strategies supported them to know where the pupils were in their learning and a better understanding of who the pupils were as individuals. Teachers were able to identify which pupils needed support much quicker, and also adapt teaching in the real time to support any misconceptions that pupils had within the learning process.

 3. **Feedback: Provide feedback that enhances learning, within the moment**: Teachers discussed that the feedback strategies supported the teachers to

understand where in the learning process the pupils were. Some teachers discussed being able to give feedback in the moment this supported the pupils on the task they were completing at the time that was individual to the pupil and not just the task.

4. **Self-assessment: Allowing pupils to take ownership of their learning, reflecting their own thinking:** Pupils in the focus groups were able to identify how self-assessment impacted on their work, self-efficacy, more positively they were engaged and focused on their work.

5. **Peer assessment: Facilitate opportunities for pupils to discuss their own work for the purpose of enhancing understanding and knowledge:** Pupils understood what a talk partner was, they were able to see that the strategy had other impacts, such as academic support and social relationships. As well as being academic resources for each other.

As the studies in Chapter 4 identified positive impacts in the classroom from the perspective of the pupils and the teachers, the results also highlight the limitation of using soft outcomes to form conclusions in economic evaluation. Therefore, the decision maker who might utilise the findings from the economic evaluation in this thesis has to decide if the softer outcomes are important within their context and represent value for money. The lack of quantifiable shadow prices for soft outcomes of those derived qualitative data, invariably limits the conclusion of value for money to individual contexts.

RQ3: What is the feasibility of using health economics approaches to evaluate a large-scale education programme in schools?

Using a Cost Consequence Methodology, it was feasible to integrate an economic aspect within the evaluation of a large-scale school improvement programme. However, this research did demonstrate that there are some challenges when applying these methods to education. For example, the PSSRU (Curtis, 2013) supports the analyst to establish costs, and these centralised costs can support more accurate costing of intervention and creates a consistency across a range of evaluations. Given that the cost of a teacher's time is not standardised, this makes calculating the opportunity cost and the business as usual a very difficult task. While within health economics, including the opportunity cost is dependent on the perspective the analysis takes. The economic frameworks being developed in education advises that opportunity cost be included.

The FAIP intervention did not yield any statistically significant differences in relation to the outcome indicators. Had there been a small effect size, further methods of health economic methods could have been performed. For example, positive differences for the intervention school would have allowed the research to conduct a cost-effectiveness analysis and calculate the incremental cost-effectiveness ratio (ICER). This would have allowed decision makers to see the cost per unit of outcome over the control group. Had differences been identified on the CHU-9D for the intervention group, the researcher could have conducted a cost-utility analysis. The CHU-9D can be converted to a QALY, which would signal to the decision maker the gains to quality of life, for the intervention. Further to calculating a QALY, decision makers in health have benchmarks for excepting treatments into the NHS (this would necessarily be relevant to education, as there is not a framework to support the acceptability of interventions in line with QALYs gained).

Nevertheless, the economic framework of CCA, was demonstrated and allows the decision maker to identify the impact that are important to their context and setting (Glouberman & Zimmerman 2002). But could limit conclusions being drawn to wider set of decision makers, who many do not identify the softer outcomes and particularly with limited quantitative data.

7.2 Considering the theory of change

This thesis used a theory of change articulated through a logic model to frame the evaluation. Figure 7.1 captures the intended outcomes/impacts designed with the stakeholders and the researcher. The outcomes/impacts were split into short, medium, and long term, and will be discussed in relation to the findings from the FAIP intervention.

Figure 7.1 1 Theory of Change (Adapted logic model)

Context	Aims	Rationale/theory of action
• Wales is undergoing a radical policy change with a new curriculum being implemented over three years (Donaldson, 2015). • Wales had come under international scrutiny with regards to the outcome on the Program for International Student Assessment (PISA). • Initial teach training, had been reviewed and there is a movement to get evidence informed practice into schools (Furlong, 2015). • National priorities require Welsh schools to reduce the attainment gap using more evidence-informed strategies and ideas while having a stronger emphasis on formative assessment (Welsh Government, 2017). • Evidence indicates that formative assessment strategies have the potential be very promising approach to h... particularly fo...	• Improve the quality of teaching standards region wide. • Reduce the variance in teaching. • Embed effective formative assessment strategies in schools. • The development of pedagogy through reflection and being open to challenge. • To have a consistent approach to formative assessment approaches across the region. • Raise standards in line with national priorities. • Reduce teachers workloads in response to national agenda. • Prepare for the curriculum, the 4 purposed and the 12 pedological	• Lead teachers attending the training in principles of effective formative assessment. • Lead teachers trial formative assessment strategies within their classrooms and reflect on the impact, adjusting and/or improving strategies as required. • Lead teachers will disseminate strategies throughout the school and to partner schools in Tier 2 and 3. • The quality of teaching and learning will improve. • A reduction in teacher workload as more effective feedback strategies are employed.

Short :
- More effective use of formative assessment strategies in schools.
- More effective teaching leading to improved engagement and levels of participation by learners.

Medium :
- More effective teaching leading to improved understanding of concepts and ideas.
- Improved outcomes.

Long :
- A sustained improvement in the quality of teaching and embedded formative assessment principles in schools.
- Sustained improvement in outcomes over time, with more learners achieving their potential.

	Outcomes/impact
• All lead teachers • Teachers will hav • Support from Gwl • Consent will be ol	**Short :** • More effective use of formative assessment strategies in schools. • More effective teaching leading to improved engagement and levels of participation by learners. **Medium :** • More effective teaching leading to improved understanding of concepts and ideas. • Improved outcomes. **Long :** • A sustained improvement in the quality of teaching and embedded formative assessment principles in schools. • Sustained improvement in outcomes over time, with more learners achieving their potential.

• KESS student
• Supervisory t
• Research stud
• Continued fin from GwE
• School engag
• Teacher enga
• Parental cons data
• GwE staff to observations

pupils
• Collect attainment data from GwE
• Statistical support for PhD

Short impacts:

- **More effective use of formative assessment strategies in schools.**

The teachers were able to trial a range of strategies within their classrooms, the observational data confirmed that the improvement offices were able to identify some of the strategies being trialed. Through interviews, teachers were able to identify the positive elements from the different strategies and the impact that they see within the classroom.

- **More effective teaching leading to improved engagement and levels of participation by learners.**

Through disseminating the formative assessment strategies, the class teachers discussed the changes in the classroom ethos, including behavior and pupils displaying more self-efficacy. The teachers discussed knowing the pupils better and where they were in the learning journey.

Medium term impacts:

- **More effective teaching leading to improved understanding of concepts and ideas**.

The focus groups identified the pupils understood the strategies that were being used. They were able to explain the rationale behind the strategies and the outcomes in relation to the work or tasks they were doing. Particularly using the success criteria and having talk partners

understood what was expected from them but also how they could use each other as resources that focuses on work or task they were completing.

- **Improved outcomes.**

Some of the teachers discussed that given the success criteria and different modes of feedback, meant that the work that was returned to them was of a better standard. While not across all participants, some teacher discussed that learners were reaching above expected levels, and two teachers discussed this was particularly evident for lower achieving pupils. Importantly, the pupils themselves discussed that they were more aware of the standard of work that was expected and how they might reach that standard.

Long term impacts:

- **A sustained improvement in the quality of teaching and embedded formative assessment principles in schools.**

Long term impacts have not been measured in this relatively short evaluation; the time scale was over one academic year. Thus said, the teacher did identify that over the course of the academic year 2018/19 they were able to reduce their workloads further research would be needed to identify if the impact was long-term.

7.3 Limitations of the thesis

7.3.1 Systematic review

Given that the searches for the systematic review were executed up until 2018, additional searches were made in (April 2023) to identify if any new research that had been published since the first search was completed.

A total of 751 articles from the original databases were identified (ProQuest Social Science Premium, Jstor, Web of Science, PsycINFO and Cochrane Library [including PubMed and Medline]). Hand searching the Centre for Benefit-Cost studies in Education resulted in three additional articles. Google Scholar now has an advanced search tool, from which there were 116 articles identified.

In total, 870 titles and abstracts were screened using the original inclusion criteria. A total of five records were sought for full screening, and three were excluded. The two articles were fully screened. The first was a cost analysis of two different digital mathematic tools and the second was the cost effectiveness of a reading intervention.

Two different digital maths tools were evaluated using secondary data. As the outcome indicators were different for each tool the analysts were unable to perform a cost-effective analysis (Hollands & Pan, 2018). While this may have excluded the article from the systematic review, given there were no attempts to integrate the cost and outcomes there is a

very detailed cost analysis where the analysts provide marginal costs per additional pupil to the school and the district. This marginal cost analysis could be very helpful for decision makers given that the digital tools were already embedded in the schools, and to help decision makers to understand the scalability of the interventions. Economies of scale are an important element of decision making in health; however, short and long scale marginal costs should also be considered (Morris et al., 2012).

The second article was a cost-effectiveness analysis of a reading intervention (Scammacca et al., 2020). The data was used from a previous study and the costing was retrospective using the 'ingredients method'. The intervention 'Promotion Adolescence' Comprehension of Text' (PACT) was compared to business as usual (BAU) the Incremental Cost-effectiveness Ratio (ICER) was conducted for a one standard deviation gain on the different outcome measure from the intervention, and sensitivity analysis was conducted on the number of students exposed to the program. The original study of PACT was small scale, so the analysts also costed the implementation of PACT in a typical school setting (Scammacca et al., 2020). Given that the original PACT study did not incorporate the training of teachers and there were some discrepancies with the class sizes and BAU, the analysis presented the cost to implement the intervention over BAU in a 'typical' school setting.

The re-running of the searches demonstrated that economic methods are being used in education.

7.3.3 Qualitative research

One limitation of the qualitative research in this study was the absence of inter coder reliability (ICR). ICR is the level of agreement between two or more coders of qualitative data (O'Connor & Joffe, 2020). While the supervisors were able to support and offer advice, having ICR could have identified any bias in the coding and theme development in the qualitative elements of the research process.

7.3.4 Research design

Using a quasi-experimental design was appropriate to evaluate the 'real world' impact of the FAIP intervention and considering the intervention had already begun. There are limitations to using this design, particularly in the trustworthiness of the findings. These limitations emanate from the non-random sampling, the internal validity from selection bias for the intervention and control group. These limitations are common among quasi-experimental designs (Schweizer et al., 2016).

The evaluation was built around the second tier of the FAIP intervention so the results can only be associated with the second tier. Given that the intervention was complex with multi-faceted components (i.e., train the trainer model and numerous strategies being implemented) this could limit the finding. As with any applied school-based research there is limited control over other programs or initiatives being deployed at the same time, leading to a difficulty attributing impact (McCall & Groark, 2010).

7.3.5 Outcome measures

The evaluation window for the FAIP was one academic year and, as discussed in Chapter 6, the timescale for an evaluation is an important consideration for the research design. For complex and universally targeted interventions like the FAIP, longitudinal data should be tracked as practice becomes embedded (Greenberg & Abenavoli, 2017; Michie et al., 2018; Murray, 2019,).Previous large-scale implementation of formative assessment had a longer evaluation timeframe, with the recommendation that this type of formative assessment implementation training ideally requires an evaluation framework spanning two academic years (Anders, et al., 2022; Speckesser, et al., 2018). This project evaluated the impact of the FAIP over only one year, so it is possible that there was insufficient time for the impact of the FAIP strategies to become embedded in classroom practice and to manifest in changes in learners' National Reading and Numeracy test results.

This research also highlighted the difficulty in measuring the softer outcomes and the lack of shadow pricing made quantifying outcome difficult. Even with the quantitative data benchmarking or identifying proxy values would have been difficult given the little attempts to do this in education.

7.3.6 Economic evaluation

Given that the perspective of this research was from the education provider, and the costs supplied to the researcher were from the education provider this could obscure the real cost. For example, the education provider may secure different rates to purchase goods or services, meaning the cost could be region/ provider specific. This could limit the generalisability of the cost to others wanting to uptake a similar course of action.

 The education provider has a flat rate for supply cover to release teachers to attend training. The education provider did not give a breakdown of how they calculated the flat rate, again could limit the generalisability to other decision makers pursuing a similar course of action. Chapter 5 details sensitivity analysis was conducted on the cost of supply cover.

This changes the per pupil cost range from the base case scenario of £67.55 to £80.59 (inflater to 2023 prices) While I was explicit in the way that the cost per hour for a teacher's time was calculated, it is open to interpretation, as others may calculate this cost differently, this is less of an issue in health economics due to the centralised cost repository.

7.4 Strengths and unique contribution of the thesis

Chapter 3 describes the first systematic review of economic evaluations in education settings. This provides important new information and context to the use of economic evaluations and identifies important recommendations that will help researchers make more effective use of economic methods in education research.

This research has identified the need to accurately cost teachers' time as part of future economic evaluations. For the education system to be able to conduct economic evaluations, it is essential that we move towards more sophisticated ways to understand costs, and teachers' time is one of the most important costs to consider. Including the on costs and the facilities cost into the average hourly cost represents a more accurate cost.

This thesis was able to demonstrate the feasibility of using CCA to improve the quality of evaluation of a large-scale, complex regional school improvement intervention.

Evaluating universal and complex interventions is difficult due to the limited control the research has on the design and implementation of the project. This thesis was able to demonstrate that working alongside the education provider and producing a logic model it was clear what the stakeholders 'intended impact was. This in turn supported the development of methods of data collection. It demonstrates the importance of working with stakeholders to build an evaluative framework.

Finally, this thesis attempts to map out what the analyst needs to consider before undertaking economic evaluations in education settings.

7.5 Areas of future research

This section will describe some important areas of future research, including work currently under way with schools in North Wales to help them make more considered judgements about the value for money of interventions and programmes used in schools.

7.5.1 Programme Budgeting Marginal Analysis (BPMA)

We have worked with a cluster of headteachers to develop a way to appraise the interventions within their schools. This follows the work previously undertaken in North Wales to appraise the evidence base for provision in schools. See Pegram et al. (2022) for the results of this work.

We used an approach called Programme Budgeting Marginal Analysis (PBMA) to support schools to understand the cost implication and impact some of the interventions were having. PBMA is a priority setting framework that is used in health to maximise patient outcomes or set priorities within the fixed budget. It is a robust and clear way to facilitate decision making with regards to resource use. This method originated in the military in the USA as a cost accounting tool, (Maijama'a & Nazri, 2018). PBMA is a method that informs decision makers and supports an ethical way of making decisions and leaves behind the "archaic and ineffective" (Mitton et al., 2011, p. 95) historical allocations or political decision making (Mitton et al., 2011). There are eight stages to PBMA and are detailed in figure 7.2. The PBMA project[23] was co-constructed with a head teacher and GwE and is an example of engaging schools in the implementation of research ideas and concepts to improve provision to support the needs in the system. Working with teachers is seen as an important element to understanding why schools might not use evidence informed practice (Hollands et al., 2021).

Figure 7. 2 Stages of PBMA, (Mitton, et al., 2011)

---
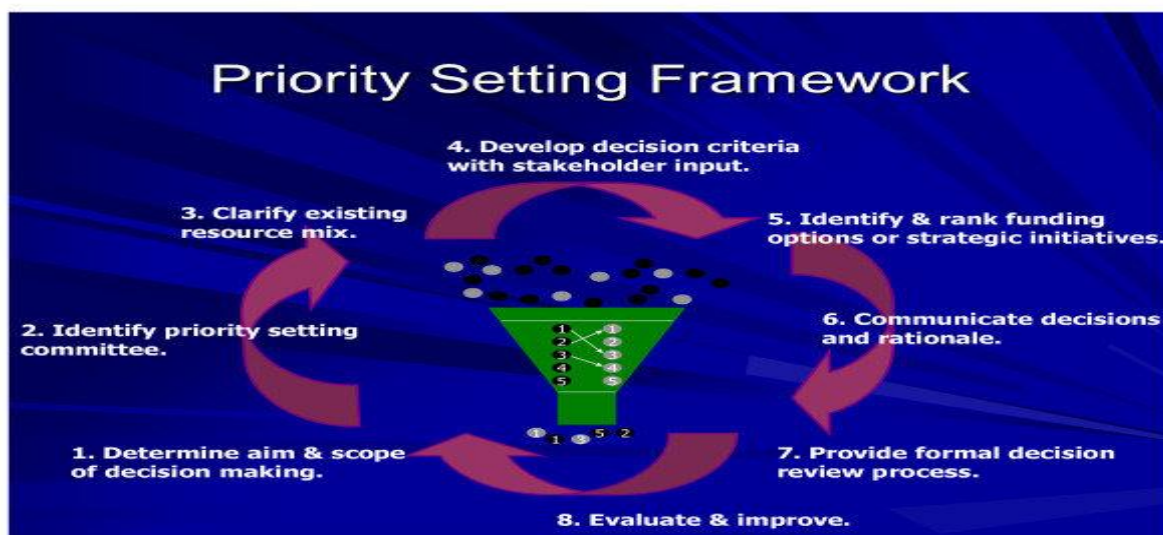
[23] The PBMA project began in the academic year 2019- 2020. However, the COVID-19 pandemic caused major disruption to collecting follow up data. Given the pressures in the system following the pandemic and stretched timescales for the researcher, with the support of the supervisory team it was decided to not continue with follow up data and the study be omitted from the thesis.

This activity involved working with a cluster of schools to appraise some of the interventions that they were running in the schools. Appendix E1 details the pro forma developed for understanding the impact, social validity and cost associated with each intervention. The work involved a group of stakeholders from each school, and the researcher supported head teachers to calculate the cost per pupil on important areas of provision in school. Each school was presented with the report with recommendations based on the information provided. See Appendix E2 for an example of one of the reports produced.

This PBMA work has been funded regionally to help schools, and light touch PBMA workshops have been delivered in six additional schools in North Wales. Emerging findings suggest that headteachers appreciated a system that supports integrating costs and outcomes, and the process was felt to be beneficial as it illuminated to staff the costs of interventions and the wider impact. Some schools were able to change provision and use the PBMA process to support school inspection activities. More in-depth research needs to be conducted on and clarify how PBMA can be used as a decision-making tool, but there are some tentative foundations being set with this work in the North Wales region.

7.5.2 Evaluation direction

Within health, analysis can utilise a central cost repository which supports the standardization on some costs contained in the health framework. A centralised cost repository in education could support more researchers to incorporate costs, even if they do not interrogate the cost against outcomes. Recent work from Shand and Bowden (2022) has explored the development of establishing common assumptions around collating costs for economic evaluations in education. Although the approach to common assumptions needs to be wider than costs alone, it paves the way for a more unified approach to conducting economic evaluations in education. The Centre for Benefit Cost Studies in Education (CBCSE, n.d.) have developed the *Cost-out*® tool for the U.S Department of Education to support researcher, educationalists and policy makers conduct economic evaluations. The *Cost-out*® tool is based on American prices, inflation, and discount rates; it allows users to input the necessary costs and the tool will produce a report based on the inputted data. The tool has been used in various research in education settings (Leyva et al., 2022; Scheibel & Zimmerman, 2022). It is an example of a more robust approach to collating cost information in the U.S education settings. As of yet, this has not been adapted for the UK, but it could be an important avenue to explore.

Although the 'ingredients method' is growing traction in the USA, in the UK there is relatively little use of this evaluative framework in education. As Chapters 3 and 6 described,

this method comes from a small base in America but is gaining momentum in other countries. Chapter 6 discussed some of the similarities and differences between health economics and the 'ingredients method'; and although both support the researcher to do the same cost evaluations.

The EEF publishes cost collecting guidance for all projects that they fund, and this now includes the opportunity cost (previous guidance and evaluations did not include the opportunity costs) (EEF, 2019). This means that in EEF funded trials the research team now need to collate the costs of the intervention, which should routinely incorporate the opportunity cost. The new EEF guidance states that the 'ingredients method' should be used for the evaluation and envisages the 'ingredients method' as a promising approach to being used in England. This could be important to the Welsh education system as both organisations look to forge stronger links with the work of the EEF (Welsh Government, 2022).

While it is outside of the scope of this thesis to evaluate which economic framework has greater utility (health economics or 'ingredients methods'), the findings here do suggest that there are promising options for researchers to consider. Furthermore, it is important that researchers are explicit regarding which framework they are using. With the EEF including the 'ingredients method' in their evaluation framework this could suggest the 'ingredients method' may take up more of the evaluative landscape in the UK.

7.5.3 Supporting the journey to evidence

We have previously seen how education is striving to become more evidence informed in many countries. In policy making there is the need for decision makers to use research evidence to disseminate policies in education. Additionally, there is a drive to encourage practitioners to using evidence-based provision in the classroom. What we do know so far is that there is a lack of evidence-based decision making in policy (Pellegrini & Vivanet, 2021) and in the classroom (Pegram et al., 2022).

While it is easy to suggest that provision needs to be evidence based, we need to understand the pathway to the generation of evidence and, therefore, the establishment of an evidence base. Chapter 3 recommends that economic evaluations should be built into the early stages of efficacy trials. Thornicroft et al. (2011) have discussed a framework within the medical sciences that works to establish from the inception of an idea/hypothesis to the implementation of activities. Owen et al. (2022) links this framework to education to support researchers and practitioners to use systematic frameworks to generate knowledge to support the journey to evidence. Figure 7.3 illustrates the phases. Interestingly, Owen et al. (2022)

suggests that as research moves to phase 3 and intersects with transitional block 3, schools may not adopt an intervention due to considerations of the ongoing cost associated with the programme delivery.

Figure 7.3 1 Journey to evidence. (Thornicroft et al., 2011)
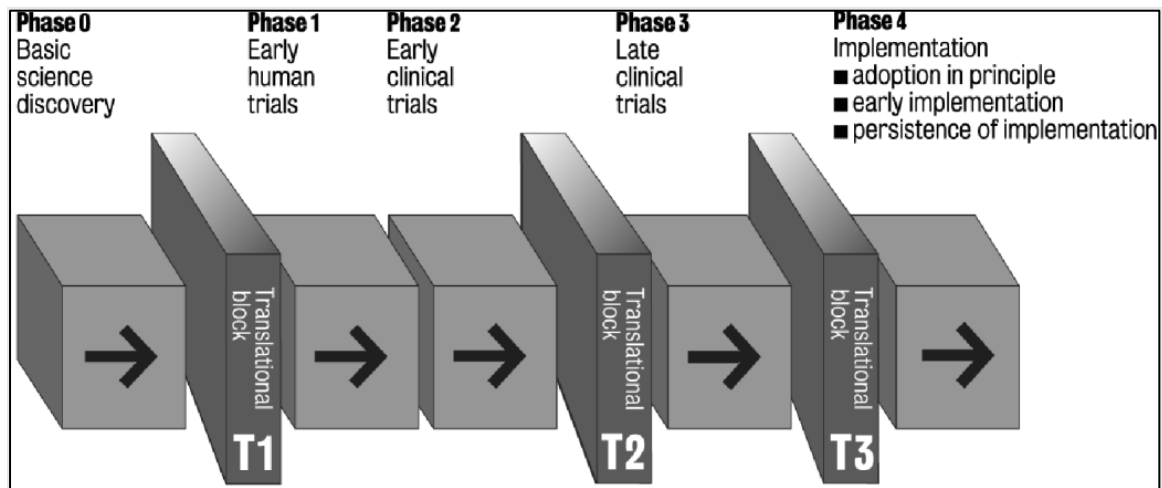
To mitigate the cost associated in transitional block 3, Owen et al. (2022) suggest that at this point researchers need to consider the economic evaluations so that costs and outcomes can be integrated alongside each other. However, if we return to health economics, costs are built around phase 0 and 1 and should be considered in transactional block one. Early identification of costs supports the scalability of interventions (Hartz & John, 2008; Ramsey et al., 2018), while Levin and Belfield (2018) and Scammacca, et al. (2020) advocate that economic evaluation should be conducted at all stages of the research to guide the decision makers around implementation. This is an important consideration for future research.

7.5.4 Measuring outcomes

While there is a push for education to consider value for money and cost-effective decisions around provision, it is also important to understand that it might be challenging to use cost-effectiveness analysis exclusively in effectiveness research in education.

The systematic review identified that there were issues with the comparability of interventions, particularly in studies where cost-effectiveness was used as the evaluation method. For a CEA to work, the approach to measuring outcomes of the comparators needs to be consistent. This is one of the main criticisms of CEA in health economics (Weintraub & Cohen, 2009), and led to the development of the QALY to try to mitigate the need for the outcome measures to be the same across studies, even when interventions were targeting

different outcomes. As discussed earlier, the QALY enabled interventions to be compared across treatments, populations and it has become a central part of decision making within health policy (MacKillop & Sheard, 2018). While there are examples of QALYs being used in education (Whiteley et al., 2022; Persson, et al., 2018) they are not designed to capture the attainment of pupils and instead focus on the health utility gained by the interventions. Given that attainment is one of the central outcomes of education, the QALY may not be a suitable measure for this field. Guidance suggests that QALYs should not be used when the main outcome is not health related (EUnetHTA, 2016).

The World Bank have developed a measure to look not only at the effect of the quantity of schooling in relation to human capital, but also the quality of schooling. This measure is called Learning-Adjusted Years of Schooling (LAYS; Filmer et al., 2018). The LAYS measure is a macro-level measure that seeks to compare across countries the difference in schooling. The focus of the measure has been at low to middle income countries, while this measure is still being developed it could support policy makers to identify country differences that other measures fail to capture. This is an interesting development for comparative education looking at macro level systems, but for the context of using it in the types of economic evaluation at the meso and micro levels, it does little to mitigate a common measure to make actions in education more comparable particularly in cost-effectiveness analysis (Glandon et al., 2023).

While it is outside of the scope of this thesis, it is important to consider a creation of a unified measure of education effectiveness that would be transferable to any intervention or courses of action, that would mitigate the need for the same outcome measure in cost-effectiveness analysis but could also measure other benefits and not reduce an intervention to a single outcome. If we look at cost utility analysis, the approach is to use outcome measures to determine individuals' health status, and then apply societally derived preference weighting to that health status to determine what the incremental utility gained from the intervention is compared to individuals in a comparator group. While an 'Education QALY' would be a suitable course of action it would be very difficult to understand the preferences around utility and costs (Levin et al., 2017). Gilead (2014) discusses the difficulties of quantifying costs and outcomes into a single unit, and some things (costs and benefits) are incommensurable. Gilead's(2014, 2018) discussion centres on the use of CBA and CUA in education and how there is still some work to do to estimate the costs (what goes in) and the different outcomes (what comes out), and until this is developed economic evaluations will only be partial ones.

7.5.5 Welfare economics

As stated in chapter 6, education has used cost benefit analysis where the utility resulting from an intervention is represented in monetary values. This has been used in education as the utility is difficult to define, and it is possibly easier for decision makers to understand utility in monetary terms (Gilead, 2012; 2018). In health economics the theory of welfare economics has supported this development of understanding utility in a much more sophisticated way. Welfare economics, like in education, uses monetary value as a proxy for utility (Gilead, 2014; Morris et al., 2012). Using the Pareto principle has helped health economists make value judgments on how best to allocate scarce resources, but within education we have yet to build an ethical framework that, firstly understands what utility is in education and secondary develop a framework to make judgments about societal preferences for various education utility states. Gilead (2014, p. 383) makes the point *"...educational investment should be based not on the basis of efficiency but rather on the considerations, such as what we ought to invest in."* This is no different to what welfare economics is trying to do; as yet education has yet to understand utility in education terms and there has been limited attempts to conceptualise utility outside of a monetary value (Gilead, 2012; 2014; 2018).

7.6 Summary

The four studies in this thesis aimed to understand and test the feasibility of using health economics in and educations setting. Study 1, the systematic review identified that there is some work in education that is building economic frameworks into the evaluation landscape, but some of the applications of the methods could limit the scope to decision makers. Of the high-quality studies included, the discipline of education is building an economic framework that is specific to education and differs in methodology to health economics.

The data from Chapter 4, (Study 2 and 3) demonstrated the impact of the FAIP intervention, and this data was able to be integrated into a wider evaluative framework in Chapter 5. The economic evaluation carried out in Chapter 5, demonstrated a health economic methodology can be used to evaluate an activity in education and draws on the limitations with regards to valuating softer outcomes in an economic framework. This thesis was able to calculate the opportunity cost of a teacher's time and is one of the first attempts to develop this costing methodology in Wales. Chapter 6 was able to map out the consideration to undertake economic evaluations in education.

Finally, some theoretical consideration needs to be developed further if the impact of education is going to be realised, where there is a deeper understanding of what the impacts means, to who and how we might measure that impact.

7.7 Reflections

The journey through this PhD, has been enjoyable and illuminating. Coming from the health perspective I quickly realised that for education and particularly education economics there is a very different landscape. All the pupils, school staff, school improvement officers, local authority officers and government representatives I have met along the way have taught me so much about *how* the education system works. It's complex, individual to each school, teacher, and pupil and our job is to support the complex, individual nature of the education system for the benefit of all pupils.

References

Adamma, O. N., Ekwutosim, O. P., & Unamba, E. C. (2018). Influence of Extrinsic and Intrinsic Motivation on Pupils Academic Performance in Mathematics. Online Submission, 2(2), 52-59.

Adams, W. C. (2015). Conducting semi-structured interviews. Handbook of practical program evaluation, 492-505.

Agasisti, T. (2014). The efficiency of public spending on education: An empirical comparison of EU countries. *European Journal of Education*, *49*(4), 543-557.

Ahmed, F. (2020). Is the Attainment Gap Fundamentally Flawed? Challenges and Opportunities. *Research in Teacher Education*, *10*(1), 18-22.

Ander, R., Guryan, J., & Ludwig, J. (2016). Improving academic outcomes for disadvantaged students: Scaling up individualized tutorials. *The Hamilton Project–Brookings*.

Anders, J., Foliano, F., Bursnall, M., Dorsett, R., Hudson, N., Runge, J., & Speckesser, S. (2022). The effect of embedding formative assessment on pupil attainment. *Journal of Research on Educational Effectiveness*, *15*(4), 748-779.

Andersson, H. (2018). Application of BCA in Europe-Experiences and Challenges. In Journal of Benefit-Cost Analysis (Vol. 9, Issue 1).

Andrews, J., Robinson, D., & Hutchinson, J. (2017). Closing the Gap?: trends in educational attainment and disadvantage. August 2017.

Auriacombe, C. J. (2011). Role of theories of change and programme logic models in policy evaluation.

Ball, S. J. (2016). Neoliberal education? Confronting the slouching beast. Policy Futures in Education, 14(8), 1046-1059.

Banerjee, P. A. (2016). A systematic review of factors linked to poor academic performance of disadvantaged students in science and maths in schools. *Cogent Education*, *3*(1), 1178441.

Barnyak, N. C., & McNelly, T. A. (2009). An Urban School District's Parent Involvement: A Study of Teachers' and Administrators' Beliefs and Practices. *School Community Journal*, *19*(1), 33-58.

Barone, C. (2019). Towards an education-based meritocracy. *ISA eSymposium for Sociology*, *9*(1), 1-8.

Barrett, J., Gonsalvez, C. J., & Shires, A. (2020). Evidence-based practice within supervision during psychology practitioner training: A systematic review. *Clinical Psychologist*, *24*(1), 3-17.

Belfield, C., Bowden, A. B., Klapp, A., Levin, H., Shand, R., & Zander, S. (2015). The Economic Value of Social and Emotional Learning. *Journal of Benefit-Cost Analysis*, *6*(3), 508–544. https://doi.org/10.1017/bca.2015.55

Bell, A. (2007). Designing and testing questionnaires for children. Journal of Research in Nursing, 12(5), 461-469.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in education: principles, policy & practice*, *18*(1), 5-25.

Bilinski, A., Neumann, P., Cohen, J., Thorat, T., McDaniel, K., & Salomon, J. A. (2017). When cost-effective interventions are unaffordable: Integrating cost-effectiveness and budget impact in priority setting for global health programs. PLoS Medicine, 14(10). https://doi.org/10.1371/journal.pmed.1002397

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, *5*(1), 7-74.

Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British educational research journal*, *29*(5), 623-637.

Black, P., & Wiliam, D. (2005). Changing teaching through formative assessment: Research and practice. *Formative assessment: Improving learning in secondary classrooms*, 223-240.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5-31.

Blamey, A., & Mackenzie, M. (2007). Theories of change and realistic evaluation: peas in a pod or apples and oranges?. *Evaluation*, *13*(4), 439-455.

Blanca Mena, M. J., Alarcón Postigo, R., Arnau Gras, J., Bono Cabré, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option?. Psicothema, 2017, vol. 29, num. 4, p. 552-557.

Borman, G. D., & Hewes, G. M. (2002). The long-term effects and cost-effectiveness of success for all. In *Educational Evaluation and Policy Analysis* (Vol. 24, Issue 4, pp. 243–266). American Educational Research Association. https://doi.org/10.3102/01623737024004243

Bowden, A. B., & Belfield, C. (2015). Evaluating the Talent Search TRIO program: A Benefit-Cost Analysis and Cost-Effectiveness Analysis1. *Journal of Benefit-Cost Analysis*, *6*(3), 572-602.

Bowden, A. B., Shand, R., Belfield, C. R., Wang, A., & Levin, H. M. (2017). Evaluating Educational Interventions That Induce Service Receipt: A Case Study Application of City Connects. *Journals.Sagepub.Com*, *38*(3), 405–419. https://doi.org/10.1177/1098214016664983

Bowden, A. B., Shand, R., Levin, H. M., Muroga, A., & Wang, A. (2020). An economic evaluation of the costs and benefits of providing comprehensive supports to students in elementary school. Prevention Science, 21, 1126-1135.

Braun, V., & Clarke, V. (2012). *Thematic analysis*. American Psychological Association.

Broadfoot, Patricia & Daugherty, Richard & Gardner, John & Gipps, Caroline & Harlen, Wynne & James, Mary & Stobart, Gordon. (1999). Assessment for Learning: beyond the black box.. 10.13140/2.1.2840.1444.

Brouwer, W. B., Culyer, A. J., van Exel, N. J. A., & Rutten, F. F. (2008). Welfarism vs. extra-welfarism. *Journal of health economics*, *27*(2), 325-338.

Brown, M., Mcnamara, G., O'brien, S., Skerritt, C., O'hara, J., Faddar, J., Cinqir, S., Vanhoof, J., Figueiredo, M., & Kurum, G. (2020). Improving Schools Parent and student voice in evaluation and planning in schools. *Journals.Sagepub.Com*, *23*(1), 85–102. https://doi.org/10.1177/1365480219895167

Bryman, A. (2016). *Social research methods*. Oxford university press.

Buxton, M. (2011). The payback of 'Payback': challenges in assessing research impact. Research Evaluation, 20(3), 259-260.

Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. Quality & quantity, 56(3), 1391-1412.

Caldwell, B. J. (2018). School leadership and professional autonomy. *Wales Journal of Education*, *20*(2).

Campbell, D. T., & Stanley, J. C. (2015). *Experimental and quasi-experimental designs for research*. Ravenio Books.

Canaway, A., Frew, E., Lancashire, E., Pallan, M., Hemming, K., Adab, P., & WAVES trial investigators. (2019). Economic evaluation of a childhood obesity prevention programme for children: Results from the WAVES cluster randomised controlled trial conducted in schools. PloS one, 14(7), e0219500.

Cardim-Dias, J. and Sibieta, L. (2022). Inequalities in Wales and England. Available at: https://epi.org.uk/wp-content/uploads/2022/07/Inequalities-in-Wales-and-England.pdf

Carrier, N. (2017). Educational Research How educational ideas catch on: the promotion of popular education innovations and the role of evidence How educational ideas catch on: the promotion of popular education innovations and the role of evidence. *Educational REsEaRch*, *59*(2), 228–240. https://doi.org/10.1080/00131881.2017.1310418

CBCSE, (N.D.) Centre for Benefit Cost Studies in Education. Available at: https://www.cbcse.org/costout

Chapko, M. K., Liu, C. F., Perkins, M., Li, Y. F., Fortney, J. C., & Maciejewski, M. L. (2009). Equivalence of two healthcare costing methods: bottom-up and top-down. *Health economics*, *18*(10), 1188-1201.

Charles, J. M., Harrington, D. M., Davies, M. J., Edwardson, C. L., Gorely, T., Bodicoat, D. H., ... & Edwards, R. T. (2019). Micro-costing and a cost-consequence analysis of the 'Girls Active'programme: A cluster randomised controlled trial. PloS one, 14(8), e0221276.

Charles, J.; Edwards, R.T. A Guide to Health Economics for Those Working in Public Health; Bangor

Chen, H. T. (2006). A theory-driven evaluation perspective on mixed methods research. *Research in the Schools*, *13*(1), 75-83.

Chen, H. T. (2006). A theory-driven evaluation perspective on mixed methods research. *Research in the Schools*, *13*(1), 75-83.

Clarke, S. (2014). *Outstanding formative assessment: Culture and practice*. Hachette UK.

Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences*, *113*(31), 8664-8668.

Clune, W. H. (2002). Methodological strength and policy usefulness of cost-effectiveness research. Cost-effectiveness and educational policy, 55-68.

Coast, J. (2004). Is economic evaluation in touch with society's health values?. Bmj, 329(7476), 1233-1236.

Collins, M., BMJ, N. L.-, & 2013, undefined. (n.d.). NICE's end of life decision making scheme: impact on population health. *Bmj.Com*. Retrieved August 19, 2020, from https://www.bmj.com/content/346/bmj.f1363.pdf+html

Colman, A. (2015). Rosenthal effect. In A Dictionary of Psychology. : Oxford University Press. Retrieved 24 Sep. 2020, from https://www-oxfordreference-com.ezproxy.bangor.ac.uk/view/10.1093/acref/9780199657681.001.0001/acref-9780199657681-e-7274.

Columbia University's Mailman School of Public Health. (2007, October 17). Reducing Class Size May Be More Cost-effective Than Most Medical Interventions. ScienceDaily. Retrieved August 1, 2021 from www.sciencedaily.com/releases/2007/10/071016131350.htm

Connell, J. P., & Kubisch, A. C. (1998). Applying a theory of change approach to the evaluation of comprehensive community initiatives: progress, prospects, and problems. New approaches to evaluating community initiatives, 2(15-44), 1-16.

Copeland, J. (2019). A critical reflection on the reasoning behind, and effectiveness of, the application of the Pupil Premium Grant within primary schools. *Management in Education*, *33*(2), 70–76. https://doi.org/10.1177/0892020618781234

Corbacho, B., & Pinto-Prades, J. L. (2012). Health economic decision-making: a comparison between UK and Spain. *British medical bulletin*, *103*(1), 5-20. Buxton, M. J. (2006). Economic evaluation and decision making in the UK. *Pharmacoeconomics*, *24*, 1133-1142.

Cost Analysis: A Toolkit (IES 2020-001). U.S. Department of Education. Washington, DC: Institute of Education Sciences. Retrieved [month day, year] from https://ies.ed.gov/pubsearch/.

Cowie, B., Harrison, C., & Willis, J. (2018). Supporting teacher responsiveness in assessment for learning through disciplined noticing. *The Curriculum Journal*, *29*(4), 464-478.

Crenna-Jennings, W. (2018). Key drivers of the disadvantage gap, Education in England: Annual Report 2018. London: Education Policy Institute. Available at: https://epi.org.uk/wp-content/uploads/2018/07/EPI-Annual-Report-2018-Lit-review.pdf

Crichton, H., & McDaid, A. (2016). Learning intentions and success criteria: learners' and teachers' views. The Curriculum Journal, 27(2), 190-203.

Cubi-Molla, P., Buxton, M., & Devlin, N. (2021). Allocating public spending efficiently: is there a need for a better mechanism to inform decisions in the UK and elsewhere?. Applied Health Economics and Health Policy, 19(5), 635-644.

Curtis, L. A. (2013). Unit costs of health and social care 2013. Personal Social Services Research Unit, University of Kent.

Darmody, M., Smyth, E., & Russell, H. (2020). The implications of the COVID-19 pandemic for policy in relation to children and young people: a research review.

Dawadi, S., Giri, R. A., & Simkhada, P. (2020). Impact of COVID-19 on the Education Sector in Nepal: Challenges and Coping Strategies. *Online Submission*.

Detrich, R. (2020). Cost-effectiveness analysis: A component of evidence-based education. *School Psychology Review*, *49*(4), 423-430.

Dickson, M., & Harmon, C. (2011). Economic returns to education: what we know, what we don't know, and where we are going—some brief pointers. *Economics of education review*, *30*(6), 1118-1122.

Donaldson, G, (2015). Successful Futures Independent Review of Curriculum and Assessment Arrangements in Wales Cardiff: Welsh Government

Doyle, L., Brady, A. M., & Byrne, G. (2016). An overview of mixed methods research–revisited. Journal of research in nursing, 21(8), 623-635.

Drayton, E., Farquharson, C., Ogden, K., Sibieta, L., Tahir, I., & Waltmann, B. (2022). Annual report on education spending in England: 2022. *The Institute for Fiscal Studies*.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*. Oxford university press.

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education. *Practical Assessment, Research, and Evaluation*, *14*(1), 7.

Dunphy, E. (2010). Assessing early learning through formative assessment: Key issues and considerations. *Irish Educational Studies*, *29*(1), 41-56.

Dweck, C. S. (2009). Mindsets: Developing talent through a growth mindset. *Olympic Coach*, *21*(1), 4-7.

Eccles, M., Grimshaw, J., Campbell, M., & Ramsay, C. (2003). Research designs for studies evaluating the effectiveness of change and improvement strategies. *BMJ Quality & Safety*, *12*(1), 47-52.

Education Scotland (2022). Pupil Equity Funding: Looking inwards, outwards, forwards – sharing effective practice to maximise support for learners and practitioners. https://education.gov.scot/media/vxxkf3pa/pupil-equity-funding-looking-inwards-outwards-forwards3.pdf

Edwards, R. T., & McIntosh, E. (Eds.). (2019). *Applied health economics for public health practice and research*. Oxford University Press.

EEF, (2018). The Attainment Gap. Available at: https://educationendowmentfoundation.org.uk/public/files/Annual_Reports/EEF_Attainment_Gap_Report_2018.pdf

EEF, (2019). Cost evaluation guidance for EEF evaluations. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/Cost_Evaluation_Guidance_2019.12.11.pdf

Estyn (2022). The Curriculum for Wales: How are regional consortia and local authorities supporting schools?. https://www.estyn.gov.wales/system/files/2022-03/The%20Curriculum%20for%20Wales%20-%20how%20are%20regional%20consortia%20and%20local%20authorities%20supporting%20schools_0.pdf

EUnetHTA Joint Action 2, Work Package 7, Subgroup 3, Heintz, E., Gerber-Grote, A., Ghabri, S., Hamers, F. F., Rupel, V. P., ... & Davidson, T. (2016). Is there a European view on health economic evaluations? Results from a synopsis of methodological guidelines used in the EUnetHTA partner

Evans, G. (2021). Back to the future? Reflections on three phases of education policy reform in Wales and their implications for teachers. Journal of Educational Change, 1-26.

Eyles, A., Machin, S., & McNally, S. (2017). Unexpected school reform: Academisation of primary schools in England. Journal of Public Economics, 155, 108-121.

Farquharson, C., Sibieta, L., Tahir, I., & Waltman, B. (2021). *2021 annual report on education spending in England* (No. R204). IFS Report. https://ifs.org.uk/publications/15858

Fauth, B., Decristan, J., Decker, A. T., Büttner, G., Hardy, I., Klieme, E., & Kunter, M. (2019). The effects of teacher competence on student outcomes in elementary science education: The mediating role of teaching quality. *Teaching and teacher education*, *86*, 102882.

Fellows, R., & Liu, A. M. (2020). Borrowing theories: contextual and empirical considerations. Construction management and economics, 38(7), 581-588.

Ferreira, M., Martinsone, B., & Talić, S. (2020). Promoting sustainable social emotional learning at school through relationship-centered learning environment, teaching methods and formative assessment. *Journal of Teacher Education for Sustainability*, *22*(1), 21-36.

Fielding, N. G. (2012). Triangulation and mixed methods designs: Data integration with new research technologies. *Journal of mixed methods research*, *6*(2), 124-136.

Filmer, D., Rogers, H., Angrist, N., & Sabarwal, S. (2020). Learning-adjusted years of schooling (LAYS): Defining a new macro measure of education. *Economics of Education Review*, *77*, 101971.

Fleming, T. R. (2011). Addressing missing data in clinical trials. *Annals of internal medicine*, *154*(2), 113-117.

Fotaki, M. (2010). Why do public policies fail so often? Exploring health policy-making as an imaginary and symbolic construction. Organization, 17(6), 703-720.

Fotaki, M. (2010). Why do public policies fail so often? Exploring health policy-making as an imaginary and symbolic construction. *Organization*, *17*(6), 703-720.

Francis, B., Taylor, B., & Tereshchenko, A. (2019). *Reassessing'ability'grouping: improving practice for equity and attainment*. Routledge.

Frew, E., & Breheny, K. (2020). Health economics methods for public health resource allocation: a qualitative interview study of decision makers from an English local authority. Health Economics, Policy and Law, 15(1), 128-140.

Fuchs-Schündeln, N., Krueger, D., Ludwig, A., & Popova, I. (2022). The long-term distributional and welfare effects of Covid-19 school closures. *The Economic Journal*, *132*(645), 1647-1683.

Furber, G., & Segal, L. (2015). The validity of the Child Health Utility instrument (CHU9D) as a routine outcome measure for use in child and adolescent mental health services. Health and quality of life outcomes, 13(1), 1-14.

Furlong, J., Griffiths, J., Hannigan-Davies, C., Harris, A., & Jones, M. (2021). The reform of initial teacher education in Wales: from vision to reality. *Oxford Review of Education*, *47*(1), 61-78.).

Gale, T., Mills, C., & Cross, R. (2017). Socially inclusive teaching: Belief, design, action as pedagogic work. *Journal of teacher education*, *68*(3), 345-356.

Ghotra, S., McIsaac, J. L. D., Kirk, S. F., & Kuhle, S. (2016). Validation of the "Quality of Life in School" instrument in Canadian elementary school students. *PeerJ*, *4*, e1567.

Gibbons, S., McNally, S., & Viarengo, M. (2018). Does additional spending help urban schools? An evaluation using boundary discontinuities. Journal of the European Economic Association, 16(5), 1618-1668.

Gibson, F. (2007). Conducting focus groups with children and young people: strategies for success. Journal of research in nursing, 12(5), 473-483.

Gilead, T. (2012). Education and the logic of economic progress. *Journal of Philosophy of Education*, *46*(1), 113-131.

Gilead, T. (2014). Education and the rationale of cost–benefit analysis. *British Journal of Educational Studies*, *62*(4), 373-391.

Gilead, T. (2018). From allocative to technical efficiency: reconsidering the basic assumptions of educational productivity. *Journal of Educational Administration and History*, *50*(4), 270-283.

Gillard, D. (2018). Education in England: a history. *Scotland: Queen's Printer For Scotland*.

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: key concepts, approaches, and applications. Research in Social and Administrative Pharmacy, 9(3), 330-338.

Glandon, D., Fishman, S., Tulloch, C., Bhula, R., Morgan, G., Hirji, S., & Brown, L. (2023). The State of Cost-Effectiveness Guidance: Ten Best Resources for CEA in Impact Evaluations. *Journal of Development Effectiveness*, *15*(1), 5-16.

Glouberman, S. and Zimmerman, B., 2002. Complicated and complex systems: what would successful reform of Medicare look like?.

Gorard, S. (2022). Segregation and the attainment gap for permanently disadvantaged pupils in England. Educational review, 1-24.

Gorard, S., See, B. H., & Siddiqui, N. (2020). What is the evidence on the best way to get evidence into use in education? *Review of Education*, *8*(2), 570–610. https://doi.org/10.1002/rev3.3200

Gorard, S., Siddiqui, N., & See, B. H. (2022). Assessing the impact of Pupil Premium funding on primary school segregation and attainment. Research Papers in Education, 37(6), 992-1019.

Gov.uk. (2021). Pupil Premium: conditions of grant 2021 to 2022 for academies and free schools. Available at: https://www.gov.uk/government/publications/pupil-premium-allocations-and-conditions-of-grant-2021-to-2022/pupil-premium-conditions-of-grant-2021-to-2022-for-academies-and-free-schools

Gov.UK. (2023). What Works Networks. Available at: https://www.gov.uk/guidance/what-works-network

Gray, E., McCambridge, J., & Strang, J. (2005). The effectiveness of motivational interviewing delivered by youth workers in reducing drinking, cigarette and cannabis smoking among young people: Quasi-experimental pilot study. *Alcohol and Alcoholism*, *40*(6), 535-539.

Green, (2022). Private Schools and Inequality. Available at: https://ifs.org.uk/inequality/wp-content/uploads/2022/08/Private-schools-and-inequality.pdf

Greenberg, M. T., & Abenavoli, R. (2017). Universal interventions: Fully exploring their impacts and potential to produce population-level impacts. Journal of Research on Educational Effectiveness, 10(1), 40-67.

Grossoehme, D. H. (2014). Overview of qualitative research. *Journal of health care chaplaincy*, *20*(3), 109-122.

Hadden, I. R., Easterbrook, M. J., Nieuwenhuis, M., Fox, K. J., & Dolan, P. (2020). Self-affirmation reduces the socioeconomic attainment gap in schools in England. *British Journal of Educational Psychology*, *90*(2), 517-536.

Haghparast-Bidgoli, H., Skordis, J., Harris-Fry, H., Krishnan, S., O'Hearn, M., Kumar, A., ... & Kadiyala, S. (2019). Protocol for the cost-consequence and equity impact analyses of a cluster randomised controlled trial comparing three variants of a nutrition-sensitive agricultural extension intervention to improve maternal and child dietary diversity and nutritional status in rural Odisha, India (UPAVAN trial). Trials, 20(1), 1-11.

Hakim, C. (2012). Research design: Successful designs for social economics research. Routledge.

Hall, R., & Pulsford, M. (2019). Neoliberalism and primary education: Impacts of neoliberal policy on the lived experiences of primary school communities. *Power and Education*, *11*(3), 241-251.

Handley, M. A., Lyles, C. R., McCulloch, C., & Cattamanchi, A. (2018). Selecting and improving quasi-experimental designs in effectiveness and implementation research. Annual review of public health, 39, 5-25.

Harden, J (2019). The True Cost of £ducation. Available at: https://www.ascl.org.uk/Our-view/Campaigns/The-True-Cost-of-Education

Harmey, S., & Moss, G. (2021). Learning disruption or learning loss: using evidence from unplanned closures to inform returning to school after COVID-19. *Educational Review*, 1-20.

Harrington, R., Taylor, G., Hollinghurst, S., Reed, M., Kay, H., & Wood, V. A. (2010). A community-based exercise and education scheme for stroke survivors: a randomized controlled trial and economic evaluation. *Clinical rehabilitation*, *24*(1), 3-15.

Harris, A. (2009). Big change question: does politics help or hinder education change?. *Journal of Educational Change*, *10*(1), 63-67.

Hartfiel, N., & Edwards, R. T. (2019). Cost–consequence analysis of public health interventions. *Applied Health Economics for Public Health Practice and Research*, 233-247.

Hartz, S., & John, J. (2008). Contribution of economic evaluation to decision making in early phases of product development: a methodological and empirical review. *International Journal of Technology Assessment in Health Care*, *24*(4), 465-472.

Hattie, J., & Clarke, S. (2018). *Visible learning: feedback*. Routledge.

Heckman, J. J. (2012). Invest in early childhood development: Reduce deficits, strengthen the economy. *The Heckman Equation*, *7*(1-2).

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of public Economics*, *94*(1-2), 114-128.

Henrikson, N. B., & Skelly, A. C. (2013). Economic Studies Part 2: Evaluating the Quality. *Evid Based Spine Care J*, *4*, 2–5. https://doi.org/10.1055/s-0033-1341594

Higgins, J., August, S. G.-A., & 2011, undefined. (n.d.). . *The Cochrane Collaboration, 2011. Available from www*.

Hilton, G. L. S. (2017). Government Policy in England on the Financing of ITT: Value for Money or a Waste of Resources? ERIC. https://eric.ed.gov/?id=ED574209

Hing, L. S., Bobocel, D. R., Zanna, M. P., Garcia, D. M., Gee, S. S., & Orazietti, K. (2011). The merit of meritocracy. *Journal of personality and social psychology*, *101*(3), 433.

Hoffmann, C., Stoykova, B. A., Nixon, J., Glanville, J. M., Misso, K., & Drummond, M. F. (2002). Do health-care decision makers find economic evaluations useful? The findings of focus group research in UK health authorities. Value in Health, 5(2), 71-78.

Hollands, F. M., Kieffer, M. J., Shand, R., Pan, Y., Cheng, H., & Levin, H. M. (2016). Cost-Effectiveness Analysis of Early Reading Programs: A Demonstration With Recommendations for Future Research. *Journal of Research on Educational Effectiveness*, *9*(1), 30–53. https://doi.org/10.1080/19345747.2015.1055639

Hollands, F. M., Pan, Y., Kieffer, M. J., Holmes, V. R., Wang, Y., Escueta, M., ... & Muroga, A. (2021). Comparing evidence on the effectiveness of reading resources from expert ratings, practitioner judgements, and research repositories. *Evidence & Policy*, 1-20

Hollands, F., & Pan, Y. (2018). Evaluating digital math tools in the field.

Hollands, F., Bowden, A. B., Belfield, C., Levin, H. M., Cheng, H., Shand, R., Pan, Y., & Hanisch-Cerda, B. (2014). Cost-Effectiveness Analysis in Practice: Interventions to Improve High School Completion. *Journals.Sagepub.Com*, *36*(3), 307–326. https://doi.org/10.3102/0162373713511850

Hondrich, A. L., Hertel, S., Adl-Amini, K., & Klieme, E. (2016). Implementing curriculum-embedded formative assessment in primary school science classrooms. Assessment in Education: Principles, Policy & Practice, 23(3), 353-376.

Hummel-Rossi, B., & Ashdown, J. (2002). The state of cost-benefit and cost-effectiveness analyses in education. Review of Educational Research, 72(1), 1–30. https://doi.org/10.3102/00346543072001001

Hunt, E., Tuckett, S., Robinson, D., Hutchinson, J., Coleman, C. (2021). COVID-19 and Disadvantage in England 2020. Available at: https://epi.org.uk/wp-content/uploads/2022/02/EPI-Disadvantage_Gaps_in_England_2022.pdf

Hursh, D. (2005). Neo-liberalism, markets and accountability: Transforming education and undermining democracy in the United States and England. *Policy Futures in Education*, *3*(1), 3-15.

Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., Augustovski, F., Briggs, A. H., Mauskopf, J., & Loder, E. (2013). Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Cost Effectiveness and Resource Allocation*, *11*(1). https://doi.org/10.1186/1478-7547-11

Hymel, S., & Katz, J. (2019). Designing classrooms for diversity: Fostering social inclusion. Educational Psychologist, 54(4), 331-339.

Ilegbusi, M. I. (2013). An analysis of the role of rewards and punishment in motivating school learning. *Computing, Information Systems & Development Informatics*, *4*(1), 35-38.

Institute of Education Sciences. (2020). *Cost Analysis: A Toolkit* (IES 2020-001). U.S. Department of Education. Washington, DC: Institute of Education Sciences. Retrieved [month day, year] from https://ies.ed.gov/pubsearch/.

Ireson, J., Hallam, S., Hack, S., Clark, H., & Plewis, I. (2002). Ability grouping in English secondary schools: effects on attainment in English, mathematics and science. *Educational Research and Evaluation*, *8*(3), 299-318.

Isen, A. M. (2001). An influence of positive affect on decision making in complex situations: Theoretical issues with practical implications. Journal of consumer psychology, 11(2), 75-85.

Jackson, B., & Marsden, D. (2012). Education and the working class (RLE Edu L Sociology of Education). Routledge.

Jackson, C. K. (2020). Does school spending matter? The new literature on an old question. American Psychological Association.

Jackson, C. K., & Mackevicius, C. (2021). The distribution of school spending impacts (No. w28517). National Bureau of Economic Research.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology*, *17*(1), 1-10.

James, C., & Colebourne, D. (2004). Managing the performance of staff in LEAs in Wales: Practice, Problems and Possibilities. Educational Management Administration & Leadership, 32(1), 45-65.

James, M. (2011). Assessment for learning: Research and policy in the (dis) United Kingdom. In *Assessment reform in education* (pp. 15-32). Springer, Dordrecht.

James, M. (2017). Embedding formative assessment in classroom practice. In *Life in Schools and Classrooms* (pp. 509-525). Springer, Singapore.

Jayasundara, K., Krahn, M., Mamdani, M., Hoch, J. S., & Grootendorst, P. (2017). Differences in Incremental Cost-Effectiveness Ratios for Common Versus Rare Conditions: A Case from Oncology. *PharmacoEconomics - Open*, *1*(3), 167–173. https://doi.org/10.1007/s41669-017-0022-7

Jones, M., & Stanley, G. (2008). Children's lost voices: ethical issues in relation to undertaking collaborative, practice-based projects involving schools and the wider community. *Educational Action Research*, *16*(1), 31-41.

Justman, M. (2018). Randomized controlled trials informing public policy: Lessons from project STAR and class size reduction. European Journal of Political Economy, 54, 167-174.

Kaffenberger, M. (2021). Modelling the long-run learning impact of the Covid-19 learning shock: Actions to (more than) mitigate loss. *International Journal of Educational Development*, *81*, 102326.

Kallio, H., Pietilä, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. Journal of advanced nursing, 72(12), 2954-2965.

Kanjee, A., & Mthembu, J. (2015). Assessment literacy of foundation phase teachers: An exploratory study. *South African Journal of Childhood Education*, *5*(1), 142-168.

Kennedy, C. H. (1992). Trends in the measurement of social validity. The Behavior Analyst, 15, 147-156.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational measurement: Issues and practice*, *30*(4), 28-37.

Kislitsyn, D.V., Popova, E.Y. (2019). Normative Economics of Education Distribution: Equity-Efficiency Trade-off. In: Solovev, D. (eds) Smart Technologies and Innovations in Design for Control of Technological Processes and Objects: Economy and Production. FarEastCon 2018. Smart Innovation, Systems and Technologies, vol 139. Springer, Cham. https://doi.org/10.1007/978-3-030-18553-4_45

Kneen, J., Breeze, T., Thayer, E., John, V., & Davies-Barnes, S. (2021). Pioneer teachers: How far can individual teachers achieve agency within curriculum development? Journal of Educational Change, 1-22.

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kraft, M. A., & Falken, G. T. (2020). A Blueprint for Scaling Tutoring across Public Schools. EdWorkingPaper No. 20-335. *Annenberg Institute for School Reform at Brown University*.

Larson, M., Cook, C. R., Fiat, A., & Lyon, A. R. (2018). Stressed teachers don't make good implementers: Examining the interplay between stress reduction and intervention fidelity. *School Mental Health*, *10*, 61-76.

Leahy, S., & Wiliam, D. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms*. Hawker Brownlow Education.

Levin, H. (2001). Waiting for godot: Cost-effectiveness analysis in education. *New Directions for Evaluation*, *2001*(90), 55–68. https://doi.org/10.1002/ev.12

Levin, H. M. (1988). Cost-Effectiveness and Educational Policy. *Educational Evaluation and Policy Analysis*, *10*(1), 51–69. https://doi.org/10.3102/01623737010001051

Levin, H. M. (2012). More than just test scores. *Prospects*, *42*, 269-284.

Levin, H. M. , & Bachman, J. G. (1972). The Effects of Dropping Out. The Costs to the Nation of Inadequate Education: A Report Prepared for the Select Committee on Equal Educational Opportunity of the. https://eric.ed.gov/?id=ED072171

Levin, H. M., & Belfield, C. (2015). Guiding the development and use of cost-effectiveness analysis in education. *Journal of research on educational effectiveness*, *8*(3), 400-418.

Levin, H. M., Glass, G. V, & Meister, G. R. (1987). Cost-Effectiveness of Computer-Assisted Instruction. *Journals.Sagepub.Com*, *11*(1), 50–72. https://doi.org/10.1177/0193841X8701100103

Levin, H., McEwan, P., Belfield, C., & Bowden, A. (2017). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis*.

Lewis, A., & Porter, J. (2007). Research and pupil voice.

Leyva, D., Shapiro, A., Yeomans-Maldonado, G., Weiland, C., & Leech, K. (2022). Positive impacts of a strengths-based family program on Latino kindergarteners' narrative language abilities. *Developmental Psychology*, *58*(5), 835

Leyva, R. (2009). No child left behind: A neoliberal repackaging of social Darwinism. *Journal for Critical Education Policy Studies*, *7*(1), 365-381.

Li, Y., & Bates, T. C. (2019). You can't change your basic ability, but you work at things, and that's how we get hard things done: Testing the role of growth mindset on response to setbacks, educational attainment, and cognitive ability. *Journal of Experimental Psychology: General*, *148*(9), 1640.

Limwattananon, S. (2014). Sensitivity analysis for handling uncertainty in an economic evaluation. Journal of the Medical Association of Thailand= Chotmaihet Thangphaet, 97, S59-64.

Lin, S. (1998). Labor income taxation and human capital accumulation. *Journal of Public Economics*, *68*(2), 291-302.

Little, A. W. (1996). Globalisation and educational research: Whose context counts?. *International Journal of Educational Development*, *16*(4), 427-438.Ustun, U., & Eryilmaz, A. (2018). Analysis of Finnish Education System to Question the Reasons behind Finnish Success in PISA. Online Submission, 2(2), 93-114

Lomas, J., Ochalek, J., & Faria, R. (2022). Avoiding opportunity cost neglect in cost-effectiveness analysis for health technology assessment. Applied Health Economics and Health Policy, 20, 13-18.

Lucas, M. (2020). External barriers affecting the successful implementation of mobile educational interventions. *Computers in Human Behavior*, *107*, 105509.

Ludwig, J., & Phillips, D. A. (2007). The benefits and costs of Head Start.

Lundkvist, A., El-Khatib, Z., Kalra, N., Pantoja, T., Leach-Kemon, K., Gapp, C., & Kuchenmüller, T. (2021). Policy-makers' views on translating burden of disease estimates in health policies: bridging the gap through data visualization. Archives of Public Health, 79(1), 1-11.

Lupton, R., & Thomson, S. (2015). WP13 The Coalition's Record on Schools: Policy, Spending and Outcomes 2010-2015 The Coalition's Record on Schools: Policy, Spending and Outcomes 2010-2015. In *sticerd.lse.ac.uk*. http://sticerd.lse.ac.uk/dps/case/spcc/wp13.pdf

Machin, S., & McNally, S. (2008). The literacy hour. *Journal of Public Economics*, *92*(5-6), 1441-1462.

Machin, S., Mcnally, S., & Wyness, G. (2013). Educational Research Educational attainment across the UK nations: performance, inequality and evidence Educational attainment across the UK nations: performance, inequality and evidence. *Educational Research*, *55*(2), 139–164. https://doi.org/10.1080/00131881.2013.801242

MacKillop, E., & Sheard, S. (2018). Quantifying life: understanding the history of quality-adjusted life-years (QALYs). *Social Science & Medicine*, *211*, 359-366.

Maijama'a, B., & Nazri, E. M. (2018). Adjusted-Program-Budget Marginal-Analysis for Student Development Activity Planning and Budget Allocation. *International Journal of Engineering & Technology*, *7*(3.20), 55-59.

Malterud , K. (2001). Qualitative research: Standards, challenges, and guidelines . The Lancet , 358 (9280), 483 – 488 .

Mauskopf, J. A., Paul, J. E., Grant, D. M., & Stergachis, A. (1998). The role of cost-consequence analysis in healthcare decision-making. In *PharmacoEconomics* (Vol. 13, Issue 3, pp. 277– 288). https://doi.org/10.2165/00019053-199813030-00002

McCall, R. B., & Groark, C. J. (2010). Challenges and issues in designing applied research. *Applied research in child and adolescent development: A practical guide*, 101-129.

McCarney, R., Warner, J., Iliffe, S., Van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: a randomised, controlled trial. *BMC medical research methodology*, *7*(1), 1-8.

McGillicuddy, D., & Devine, D. (2018). "Turned off" or "ready to fly"–Ability grouping as an act of symbolic violence in primary school. Teaching and Teacher Education, 70, 88-99.

McMillan, J. H., Venable, J. C., & Varier, D. (2013). Studies of the effect of formative assessment on student achievement: So much more is needed. Practical Assessment, Research, and Evaluation, 18(1), 2

Michie, S., West, R., Sheals, K., & Godinho, C. A. (2018). Evaluating the effectiveness of behavior change techniques in health-related behavior: a scoping review of methods used. Translational behavioral medicine, 8(2), 212-224.

Mijs, J. J. (2016). The unfulfillable promise of meritocracy: Three lessons and their implications for justice in education. *Social Justice Research*, *29*, 14-34

Mitton, C., Dionne, F., Damji, R., Campbell, D., & Bryan, S. (2011). Difficult decisions in times of constraint: criteria based resource allocation in the Vancouver Coastal Health Authority. *BMC health services research*, *11*(1), 1-11.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *Open medicine : a peer-reviewed, independent, open-access journal*, *3*(3), e123–e130.

Morris, R. and Dobson, G. (2021). Spending the pupil premium: What influences leaders' decision-making?. *Educational Management Administration & Leadership*, *49*(2), pp.284-302. DOI: https://doi.org/10.1177/1741143220905062

Morris, S., Devlin, N., Parkin, D., ed.), A. S.-E. A. in H. C. (2nd, & 2012, undefined. (n.d.). Principles of economic evaluation in health care.

Mowat, J. G. (2018). Closing the attainment gap-a realistic proposition or an elusive pipe-dream? *Journal of Education Policy*, *33*(2), 299–321. https://doi.org/10.1080/02680939.2017.1352033

Murray, N. (2021). Long-Term Impact Evaluation of the Amazonía Lee Reading Intervention in Peru. Development.

Nelson, E. (2015). Opening up to student voice: Supporting teacher learning through collaborative action research. *Learning Landscapes*, *8*(2), 285-299.

Nelson, J., & Campbell, C. (2017). Evidence-informed practice in education: Meanings and applications. *Educational Research*, *59*(2), 127-135.

Neumann, P., Jama, J. C.-, & 2018, undefined. (n.d.). QALYs in 2018—advantages and concerns. Jamanetwork.Com. Retrieved August 19, 2020, from https://jamanetwork.com/journals/jama/article-abstract/2682917

Ng, Betsy. "The neuroscience of growth mindset and intrinsic motivation." Brain sciences 8.2 (2018): 20.

Ng-Knight, T., Shelton, K. H., Riglin, L., Frederickson, N., McManus, I. C., & Rice, F. (2019). 'Best friends forever'? Friendship stability across school transition and associations with mental health and educational attainment. *British Journal of Educational Psychology*, *89*(4), 585-599.

Ní Chróinín, D., & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: teacher perspectives and experiences. *Physical Education and Sport Pedagogy*, *18*(2), 219-233.

Northern Ireland Audit Office (NIAO) (2021). Closing the Gap – Social Deprivation and links to Educational Attainment. https://www.niauditoffice.gov.uk/files/niauditoffice/media-files/249503%20NIAO%20Closing%20the%20Gap%20report%20Final%20WEB.pdf

Nthontho, M. (2017). Children as stakeholders in education: Does their voice matter?. *South African Journal of Childhood Education*, *7*(1), 1-7.

Numata, H. (2003). What children have lost by the modernisation of education: A comparison of experiences in Western Europe and Eastern Asia. *International review of education*, *49*, 241-264.

O'Connor, C. and Joffe, H. (2020). Intercoder reliability in qualitative research: debates and practical guidelines. International journal of qualitative methods, 19. DOI: https://doi.org/10.1177/1609406919899220

O'Sullivan, A., Thompson, D., Health, M. D.-V. in, & 2005, undefined. (n.d.). Collection of health-economic data alongside clinical trials: is there a future for piggyback evaluations? *Elsevier*. Retrieved August 19, 2020, from https://www.sciencedirect.com/science/article/pii/S1098301510602458

OECD (2014). Improving schools in Wales: An OECD perspective. Paris: OECD

OECD (2022), Value for Money in School Education: Smart Investments, Quality Outcomes, Equal Opportunities, OECD Publishing, Paris. https://doi.org/10.1787/f6de8710-en.

OECD, (2020). Public Spending on Education. Available at: https://data.oecd.org/eduresource/public-spending-on-education.htm

Owen, K. L., Watkins, R. C., & Hughes, J. C. (2022). From evidence-informed to evidence-based: An evidence building framework for education. *Review of Education*, *10*(1), e3342.

Ozan, C., & Kincal, R. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences-Theory & Practice*, *18*.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and policy in mental health and mental health services research*, *42*, 533-544.

Pallant, J., & Manual, S. S. (2011). A step-by-step guide to data analysis using spss 4th edition.

Palumbo, R., & Manna, R. (2019). Making educational organizations able to change: a literature review. *International Journal of Educational Management*.

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74-98.

Paterson, L., & Iannelli, C. (2007). Social class and educational attainment: A comparative study of England, Wales, and Scotland. Sociology of Education, 80(4), 330-358.

Pegram, J., Watkins, R. C., Hoerger, M., & Hughes, J. C. (2022). Assessing the range and evidence-base of interventions in a cluster of schools. *Review of Education*, *10*(1), e3336.

Pelinescu, E. (2015). The impact of human capital on economic growth. *Procedia Economics and Finance*, *22*, 184-190.

Pellegrini, M., & Vivanet, G. (2021). Evidence-based policies in education: Initiatives and challenges in Europe. *ECNU Review of Education*, *4*(1), 25-45.

Perry, J., Lundie, D., & Golder, G. (2019). Metacognition in schools: what does the literature suggest about the effectiveness of teaching metacognition in schools?. Educational Review, 71(4), 483-500.

Persson, M., Wennberg, L., Beckman, L., Salmivalli, C., & Svensson, M. (2018). The cost-effectiveness of the Kiva Antibullying Program: results from a decision-analytic model. *Prevention science*, *19*, 728-737.

Petticrew, M., Rehfuess, E., Noyes, J., Higgins, J. P., Mayhew, A., Pantoja, T., ... & Sowden, A. (2013). Synthesizing evidence on complex interventions: how meta-analytical, qualitative, and mixed-method approaches can contribute. *Journal of clinical epidemiology*, *66*(11), 1230-1243.

Ponce, O. A., & Pagán-Maldonado, N. (2017). Educational research in the 21st century: challenges and opportunities for scientific effectiveness. IJERI: International Journal of Educational Research and Innovation, (8), 24-37.

Power, S. (2016). The politics of education and the misrecognition of Wales. *Oxford Review of Education*, *42*(3), 285-298.

Punch, S. (2002). Research with children: The same or different from research with adults?. Childhood, 9(3), 321-341.

Pye, J., Mollidor, C., Taylor, C. M., & Huxley, K. (2015). Evaluation of the Pupil Deprivation Grant: 2nd Interim report, December 2015.

Quigley, A., Muijs, D., & Stringer, E. (2018). Metacognition and self-regulated learning: guidance report.

Quinn, B., Van Mondfrans, A., & Worthen, B. R. (1984). Cost-Effectiveness of Two Math Programs as Moderated by Pupil SES. *Educational Evaluation and Policy Analysis*, *6*(1), 39–52. https://doi.org/10.3102/01623737006001039.

Ramsey, S. D., Willke, R. J., Glick, H., Reed, S. D., Augustovski, F., Jonsson, B., ... & Sullivan, S. D. (2015). Cost-effectiveness analysis alongside clinical trials II—an ISPOR Good Research Practices Task Force report. *Value in health*, *18*(2), 161-172.

Rand, L. Z., & Kesselheim, A. S. (2021). Controversy over using quality-adjusted life-years in cost-effectiveness analyses: A systematic literature review. *Health Affairs, 40*(9), 1402-1410,1410A-1410I. doi:https://doi.org/10.1377/hlthaff.2021.00343

Ratnam-Lim, C. T. L., & Tan, K. H. K. (2015). Large-scale implementation of formative assessment practices in an examination-oriented culture. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 61-78.

Read, S., Macer, M. and Parfitt, A. (2020) Effective use of Pupil Premium Plus to improve educational outcomes for looked after children. DOI: https://doi.org/10.17870/bathspa.ac.uk.00013242

Rees, G., & Taylor, C. (2015). Is there a'crisis' in Welsh education. *Transactions of the Honourable Society of Cymmrodorion*, *2015*, 97-113.

Rege, M., Hanselman, P., Solli, I. F., Dweck, C. S., Ludvigsen, S., Bettinger, E., ... & Yeager, D. S. (2020). How can we inspire nations of learners? An investigation of growth mindset and challenge-seeking in two countries. American Psychologist.

Reichardt, R. E. (2000). The cost of class size reduction: Advice for policymakers. The RAND Graduate School.

Reynolds, A. J., & Temple, J. A. (2008). Cost-effective early childhood development programs from preschool to third grade. *Annu. Rev. Clin. Psychol.*, *4*, 109-139.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2002). Age 21 cost-benefit analysis of the title I chicago child-parent centers. In *Educational Evaluation and Policy Analysis* (Vol. 24, Issue 4, pp. 267–303). American Educational Research Association. https://doi.org/10.3102/01623737024004267

Reynolds, D (2015) National model for regional working (Guidance document, 193/2015) Cardiff, GB. Welsh Government

Rice, S. M. (2010). Getting our best teachers into disadvantaged schools: Differences in the professional and personal factors attracting more effective and less effective teachers to a school. *Educational Research for Policy and Practice*, *9*, 177-192.

Robertson, L. (2021). The Poverty-related Attainment Gap: A review of the evidence.

Rogers, J. C., Simmons, E. A., Convery, I., & Weatherall, A. (2008). Public perceptions of opportunities for community-based renewable energy projects. *Energy policy*, *36*(11), 4217-4226.

Rouse, C. E. (2017). The economics of education and policy: Ideas for a principles course. *The Journal of Economic Education*, *48*(3), 229-237.

Ruggiero, J. (2007). Measuring the cost of meeting minimum educational standards: An application of data envelopment analysis. *Education Economics*, *15*(1), 1-13.

Ryan, J. E. (2004). The perverse incentives of the no child left behind act. NYUL Rev., 79, 932.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54-67.

Salkeld, G., Davey, P., Policy, G. A.-H., & 1995, undefined. (n.d.). A critical review of health-related economic evaluations in Australia: implications for health policy. Elsevier. Retrieved August 19, 2020, from https://www.sciencedirect.com/science/article/pii/0168851094006725

Scammacca, N., Swanson, E., Vaughn, S., & Roberts, G. (2020). Cost-Effectiveness of a Grade 8 Intensive Reading and Content Learning Intervention. *School Psychology Review*, *49*(4), 374-385.

Scammacca, N., Swanson, E., Vaughn, S., & Roberts, G. (2020). Cost-Effectiveness of a Grade 8 Intensive Reading and Content Learning Intervention. *School Psychology Review*, *49*(4), 374-385.

Scheibel, G., Zane, T. L., & Zimmerman, K. N. (2022). An economic evaluation of emerging and ineffective interventions: Examining the role of cost when translating research into practice. *Exceptional Children*, *88*(3), 245-262..

Schleicher, A. (2018). Insights and interpretations. *Pisa 2018*, *10*.

Schleider, J., & Weisz, J. (2018). A single-session growth mindset intervention for adolescent anxiety and depression: 9-month outcomes of a randomized trial. Journal of Child Psychology and Psychiatry, 59(2), 160-170.

Schwab, S., Resch, K., & Alnahdi, G. (2021). Inclusion does not solely apply to students with disabilities: Pre-service teachers' attitudes towards inclusive schooling of all students. *International Journal of Inclusive Education*, 1-17..

Schweizer, M. L., Braun, B. I., & Milstone, A. M. (2016). Research methods in healthcare epidemiology and antimicrobial stewardship—quasi-experimental designs. *Infection control & hospital epidemiology*, *37*(10), 1135-1140.

Scott, S., Limbert, C., & Sykes, P. (2021). Work-related stress among headteachers in Wales: Prevalence, sources, and solutions. Educational Management Administration & Leadership, 17411432211054630.

Seixas, B. V. (2017). Welfarism and extra-welfarism: a critical overview. *Cadernos de saude publica*, *33*.

Shand, R., & Bowden, A. B. (2022). Empirical support for establishing common assumptions in cost research in education. Journal of Research on Educational Effectiveness, 15(1), 103-129.

Sharples, J., Albers, B., & Fraser, S. (2018). Putting evidence to work: a school's guide to implementation.

Sibieta, (2021). Comparing education catch-up spending within and outside the UK. Available at: https://epi.org.uk/publications-and-research/comparing-education-catch-up-spending-within-and-outside-the-uk/

Sibieta, L. (2020). Review of School Spending in Wales. Available at: https://www.gov.wales/sites/default/files/publications/2020-10/review-of-school-spending-in-wales.pdf

Skerritt, C. (2023). School autonomy and the surveillance of teachers. *International Journal of Leadership in Education*, *26*(4), 553-580.

Smith, E., & Gorard, S. (2005). 'They don't give us our marks': the role of formative feedback in student progress. *Assessment in Education: Principles, Policy & Practice*, *12*(1), 21-38.

Snell, J., & Lefstein, A. (2018). "Low ability," participation, and identity in dialogic pedagogy. American Educational Research Journal, 55(1), 40-78.

Solé-Auró, A., Alcañiz, M. Educational attainment, gender and health inequalities among older adults in Catalonia (Spain). Int J Equity Health 15, 126 (2016).

Speckesser, S., Runge, J., Foliano, F., Bursnall, M., Hudson-Sharp, N., Rolfe, H., & Anders, J. (2018). Embedding formative assessment: Evaluation report and executive summary.

Sternberg, R. J., Birney, D., Kirklik, A., Stemler, S., Jarvin, L., & Grigorenko, E. L. (2013). From molehill to mountain: The process of scaling up educational interventions (firsthand experience upscaling the theory of successful intelligence). In *Translating Theory and Research Into Educational Practice* (pp. 205-221). Routledge.

Stevens, K. (2012). Valuation of the child health utility 9D index. *Pharmacoeconomics*, *30*, 729-747.

Stîngu, M. M. (2012). Reflexive practice in teacher education: facts and trends. *Procedia-Social and Behavioral Sciences*, *33*, 617-621.

Sukamolson, S. (2007). Fundamentals of quantitative research. *Language Institute Chulalongkorn University*, *1*(3), 1-20.

Talbot, C. (2009). Public value—The next "big thing" in public management?.

Taylor, C. (2018). The reliability of free school meal eligibility as a measure of socio-economic disadvantage: Evidence from the millennium cohort study in Wales. *British Journal of Educational Studies*, *66*(1), 29-51.

Taylor-Robinson, D. C., Milton, B., Lloyd-Williams, F., O'Flaherty, M., & Capewell, S. (2008). Planning ahead in public health? A qualitative study of the time horizons used in public health decision-making. BMC Public Health, 8(1), 1-7.

Thornicroft. G. H., Lempp., H., & Tansella, M. (2011). The place of implementation science in the translational medicine continuum. *Psychological Medicine, 41*(11), 2015-2021. doi: 10.107/S0033291711000109

Thyer, B. A. (2012). *Quasi-experimental research designs*. Oxford University Press.

Todd, L., & Nind, M. (2011). Giving voice in educational research. *International Journal of Research & Method in Education*, *34*(2), 115-116.

Trail-Constant, T. M. (2019). Lowering the Stakes! How to reduce cheating and encourage student mastery all at the same time.

Twist, L., Jones, E. and Treleaven, O. (2022). The Impact of Covid-19 on pupil attainment. NFER: Slough.

Vaismoradi, M., Jones, J., Turunen, H., & Snelgrove, S. (2016). Theme development in qualitative content analysis and thematic analysis.

van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2020). Problematic cost–utility analysis of interventions for behavior problems in children and adolescents. New Directions for Child and Adolescent Development, 2020(172), 89-102.

Vassall, A., Sweeney, S., Kahn, J., Gomez Guillen, G., Bollinger, L., Marseille, E., ... & Levin, C. (2017). Reference case for estimating the costs of global health services and interventions.

Vella, F., & Gregory, R. G. (1996). Selection bias and human capital investment: Estimating the rates of return to education for young males. *Labour Economics*, *3*(2), 197-219.

Vetter, M., O'connor, H. T., O'dwyer, N., Chau, J., & Orr, R. (2020). 'Maths on the move': Effectiveness of physically-active lessons for learning maths and increasing physical activity in primary school students. *Journal of Science and Medicine in Sport*, *23*(8), 735-739.

Viteritti, J. P. (2010). School choice and market failure: How politics trumps economics in education and elsewhere.

Weatherly, H., Drummond, M., Claxton, K., Cookson, R., Ferguson, B., Godfrey, C., ... & Sowden, A. (2009). Methods for assessing the cost-effectiveness of public health interventions: key challenges and recommendations. Health policy, 93(2-3), 85-92.

Weintraub, N., & Bar-Haim Erez, A. (2009). Quality of life in school (QoLS) questionnaire: development and validity. *The American Journal of Occupational Therapy*, *63*(6), 724-731.

Weintraub, W. S., & Cohen, D. J. (2009). The limits of cost-effectiveness analysis. *Circulation: Cardiovascular Quality and Outcomes*, *2*(1), 55-58.

Welsh Government (2015) https://www.gov.wales/sites/default/files/publications/2018-03/teaching-tomorrow%E2%80%99s-teachers.pdf

Welsh Government (2018a). Pupils Development Grant: letter. https://gov.wales/pupil-development-grant-letter

Welsh Government (2022) https://www.gov.wales/radical-action-needed-create-truly-equitable-education-system-all

Welsh Government, (2021) The National Strategy for Educational Research and Enquiry (NSERE): vision document. Available at: https://www.gov.wales/national-strategy-educational-research-and-enquiry-nsere-vision-document-html#:~:text=priorities%20and%20domains-,Aim,enquiry%20undertaken%20by%20educational%20professionals .

Welsh Government. (2018b). Professional standards for teaching and leadership.

Welsh Government. (2019a). Draft Curriculum for Wales 2022: a guide to Curriculum for Wales 2022. Cardiff : Welsh Government

White, J., Connelly, G., Thompson, L., & Wilson, P. (2013). Assessing children's social and emotional wellbeing at school entry using the strengths and difficulties questionnaire: professional perspectives. Educ Res, 55, 87-98.

Whitehead, Sarah J., and Shehzad Ali. "Health outcomes in economic evaluation: the QALY and utilities." British medical bulletin 96, no. 1 (2010): 5-21.

Whiteley, H., Edwards, R. T., Hutchings, J., & Bowes, L. (2022). The long life-time shadow of bullying: cost-effectiveness of KiVa to reduce bullying in primary schools. *International Journal of Population Data Science*, *7*(2). https://doi.org/10.23889/ijpds.v7i2.1745

Wiliam, D. (2020). RESEARCH INTO PRACTICE. *Getting Evidence into Education: Evaluating the Routes to Policy and Practice*.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: a meta-analysis of educational feedback research. Frontiers in Psychology, 10, 3087.

Woods, C., Armstrong, P., & Pearson, D. (2012). Facilitating primary head teacher succession in England: the role of the School Business Manager. *School Leadership & Management*, *32*(2), 141-157.

Worth, J., & Van den Brande, J. (2020). Teacher Autonomy: How Does It Relate to Job Satisfaction and Retention?. *National Foundation for Educational Research*.

Wrigley, T. (2018). The power of 'evidence': Reliable science or a set of blunt tools?. *British Educational Research Journal*, *44*(3), 359-376.

WWH, (2021). What works Clearing house. Available at: https://ies.ed.gov/ncee/wwc/

Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 140-160.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... & Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. Nature, 573(7774), 364-369.

Yeh, S. S. (2007). The Cost-Effectiveness of Five Policies for Improving Student Achievement. American Journal of Evaluation, 28(4), 416–436. https://doi.org/10.1177/1098214007307928

Yeh, S. S. (2009a). Class size reduction or rapid formative assessment?: A comparison of cost-effectiveness. *Educational Research Review*, *4*(1), 7-15.

Yeh, S. S. (2009b). Shifting the bell curve: The benefits and costs of raising student achievement. Evaluation and Program Planning, 32(1), 74–82. https://doi.org/10.1016/j.evalprogplan.2008.08.006

Yeh, S. S. (2013). A reanalysis of the effects of teacher replacement using value-added modeling. *Teachers College Record*, *115*(12), 1-35

Yeh, S. S. (2020). Educational accountability, value-added modeling, and the origin of the achievement gap. *Education and Urban Society*, *52*(8), 1181-1203.

Yoon, H. S., & Templeton, T. N. (2019). The practice of listening to children: The challenges of hearing children out in an adult-regulated world. Harvard Educational Review, 89(1), 55-84,173.

Zajacova, A., & Lawrence, E. M. (2018). The Relationship between Education and Health: Reducing Disparities Through a Contextual Approach. In *Annual Review of Public Health* (Vol. 39, pp. 273–289). Annual Reviews Inc. https://doi.org/10.1146/annurev-publhealth-031816-044628

Zerbe Jr, R. O. (2007). The legal foundation of cost-benefit analysis. *Charleston L. Rev.*, *2*, 93.

Zhang, Y., & Wildemuth, B. M. (2009). Unstructured interviews. Applications of social research methods to questions in information and library science, 222-231.

Zhou, J., Wang, Y., Bu, T., Zhang, S., Chu, H., Li, J., ... & Yang, Y. (2021). Psychological impact of COVID-19 epidemic on adolescents: a large sample study in China. *Frontiers in Psychiatry*, *12*, 769697.

Zohrabi, M. (2013). Mixed Method Research: Instruments, Validity, Reliability and Reporting Findings. Theory & practice in language studies, 3(2).

Appendix

Appendix A1 Fluid approach to theme development and reflective process

Indexing

1Barriers to implementation

    1.2Resistance from staff
    1.3 Resistance from pupils
    1.4 Resistance from SLT
    1.5 Resistance from parents
    1.6 School ethos (resistance to change)
    1.7 School context
    1.8 Classroom makeup
    1.9 Resistance from teacher
    1.10 Other

2 Facilitators to implementation :

    2.1 Behaviour change (reflexive practice)
    2.2 SLT support
    2.3 Parental involvement
    2.4 Other staff involvement
    2.5 Improving strategies already implemented
    2.6 Pupils involvement
    2.7 Other

3 Action research

    3.1 Sharing ideas with other schools and teachers (f)
    3.2 Not fostering a working relationship
    3.3 Logistics
    3.4 Language
    3.5 Cluster working

4 Benefits to strategies

    4.1 Independent learning
    4.2 Focused pupil own learning
    4.3 Understanding pupils learning (minute by minute)

Emma Tiesteel
First category was deleted

Emma Tiesteel
this category was emarginated into the above.

Emma Tiesteel
This was deleted and moved to school context

Emma Tiesteel
this was a category that was after deeper inspection.

Emma Tiesteel
This was changed to other strategies

Emma Tiesteel
Deleted only one comment moved to another section

Emma Tiesteel
Deleted other school getting more support. Merged the firs column.

Emma Tiesteel
Language and cluster working were moved to logistics.

Emma Tiesteel
Move data to confidence column. Self esteem and self

Instructions

These questions ask about how you are **today**. For each question, read all the choices and decide which one is most like you **today**.

Then put a tick in the box next to it like this ☑. Only tick **one** box for each question.

Example
Today I feel quite upset so I will tick this box.

Upset
☐ I don't feel upset today
☐ I feel a little bit upset today
☐ I feel a bit upset today
☑ I feel quite upset today
☐ I feel very upset today

Now think about and answer the rest of the questions below

1. Worried
☐ I don't feel worried today
☐ I feel a little bit worried today
☐ I feel a bit worried today
☐ I feel quite worried today
☐ I feel very worried today

2. Sad
☐ I don't feel sad today
☐ I feel a little bit sad today
☐ I feel a bit sad today
☐ I feel quite sad today
☐ I feel very sad today

3. Pain
☐ I don't have any pain today
☐ I have a little bit of pain today
☐ I have a bit of pain today
☐ I have quite a lot of pain today
☐ I have a lot of pain today

© The University of Sheffield 18.01.2008

4. Tired
☐ I don't feel tired today
☐ I feel a little bit tired today
☐ I feel a bit tired today
☐ I feel quite tired today
☐ I feel very tired today

5. Annoyed
☐ I don't feel annoyed today
☐ I feel a little bit annoyed today
☐ I feel a bit annoyed today
☐ I feel quite annoyed today
☐ I feel very annoyed today

6. School Work/Homework (such as reading, writing, doing lessons)
☐ I have no problems with my schoolwork/homework today
☐ I have a few problems with my schoolwork/homework today
☐ I have some problems with my schoolwork/homework today
☐ I have many problems with my schoolwork/homework today
☐ I can't do my schoolwork/homework today

7. Sleep
☐ Last night I had no problems sleeping
☐ Last night I had a few problems sleeping
☐ Last night I had some problems sleeping
☐ Last night I had many problems sleeping
☐ Last night I couldn't sleep at all

8. Daily routine (things like eating, having a bath/shower, getting dressed)
☐ I have no problems with my daily routine today
☐ I have a few problems with my daily routine today
☐ I have some problems with my daily routine today
☐ I have many problems with my daily routine today
☐ I can't do my daily routine today

9. Able to join in activities (things like playing out with your friends, doing sports, joining in things)
☐ I can join in with any activities today
☐ I can join in with most activities today
☐ I can join in with some activities today
☐ I can join in with a few activities today
☐ I can join in with no activities today

© The University of Sheffield 18.01.2008

Strengths and Difficulties Questionnaire

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain or the item seems daft! Please give your answers on the basis of the child's behaviour over the last six months or this school year.

Child's Name .... Male/Female

Date of Birth...

	Not True	Somewhat True	Certainly True
Considerate of other people's feelings	☐	☐	☐
Restless, overactive, cannot stay still for long	☐	☐	☐
Often complains of headaches, stomach-aches or sickness	☐	☐	☐
Shares readily with other children (treats, toys, pencils etc.)	☐	☐	☐
Often has temper tantrums or hot tempers	☐	☐	☐
Rather solitary, tends to play alone	☐	☐	☐
Generally obedient, usually does what adults request	☐	☐	☐
Many worries, often seems worried	☐	☐	☐
Helpful if someone is hurt, upset or feeling ill	☐	☐	☐
Constantly fidgeting or squirming	☐	☐	☐
Has at least one good friend	☐	☐	☐
Often fights with other children or bullies them	☐	☐	☐
Often unhappy, down-hearted or tearful	☐	☐	☐
Generally liked by other children	☐	☐	☐
Easily distracted, concentration wanders	☐	☐	☐
Nervous or clingy in new situations, easily loses confidence	☐	☐	☐
Kind to younger children	☐	☐	☐
Often lies or cheats	☐	☐	☐
Picked on or bullied by other children	☐	☐	☐
Often volunteers to help others (parents, teachers, other children)	☐	☐	☐
Thinks things out before acting	☐	☐	☐
Steals from home, school or elsewhere	☐	☐	☐
Gets on better with adults than with other children	☐	☐	☐
Many fears, easily scared	☐	☐	☐
Sees tasks through to the end, good attention span	☐	☐	☐

Signature .... Date ....

Parent/Teacher/Other (please specify:)

Thank you very much for your help

Appendix A4 Revised language

Hi Dr Erez,

Hope that you are keeping well ?

I have spoken to my supervisors about using the QoLS questionnaire, and they are happy for me to proceed. However there are a few queries with the wording of the question's. This is mainly because of cultural differences. We have discussed the wording and would like to suggest changing the following subject to your approval:

Question 1: My school looks nice ?
Question 2 : I enjoy the different social activities at school (such as break time, trips, assembly)
Question 4 : I like my class teacher (We do not have home room teachers in primary schools)
Question 10 : My classroom is nice ?
Question 22 : I am satisfied with my marks (We do not use the word grades)

The age range that we are working with is between 8-11, both my supervisors and myself thought that ' aesthetic and attractive ' was not within the range of the pupils understanding. These changes will not affect the translation into Welsh.

Hopefully this will be acceptable, so we can continue to use the questionnaire.

Kind regards,

Emma Tiesteel

Re: QoLS

Naomi weintraub <naomiweintraub@gmail.com>
To   Asnat Erez
Cc   Emma Tiesteel

↩ Reply   ↩ Reply All   → Forward

Fri 18/05/2018 16:27

(i) You replied to this message on 18/05/2018 19:58.
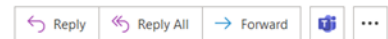
Hello Emma
Your suggestions for change are acceptable.
Good luck.
Naomi

•

פרופ' נעמי וינטראוב
ראש המעבדה לחקר ליקויים ניריו-התפתחותיים ותפקודי כתיבה
בית הספר לריפוי בעיסוק
של הדסה והאוניברסיטה העברית

Naomi Weintraub, OTR .Prof
Neuro-Developmental Disabilities and Writing Research Laboratory Head of the
School of Occupational Therapy
of Hadassah and the Hebrew University
Mt. Scopus
P.O.Box 24026
Jerusalem, Israel 9124001
naomi.weintraub@mail.huji.ac.il
5845317-2-972+ .Tel
5325345-2-972+ .Fax

Appendix A5 Quality of Life of School Children Questionnaire

No.	Statement	4 always true	3 usually, true	2 Usually not true	1 never true
1	My school looks nice	4	3	2	1
2	I enjoy different social activities at school (Break time, trips, assembly)	4	3	2	1
3	The children in my classroom treat me with respect	4	3	2	1
4	I like my class teacher	4	3	2	1
5	The subjects I learn in school are interesting	4	3	2	1
6	The walk/ ride to school is nice /comfortable	4	3	2	1
7	The chairs and desks in my school are nice/ comfortable	4	3	2	1
8	My teacher understands me	4	3	2	1
9	I feel safe in school (nothing bad will happen to me)	4	3	2	1
10	My classroom is nice	4	3	2	1
11	I have friends at school	4	3	2	1
12	I can go to my teacher with any problem	4	3	2	1
13	I feel rejected by the children in my class	4	3	2	1
14	My teachers make me feel good in class	4	3	2	1
15	My school is clean	4	3	2	1
16	The teachers in my school are good	4	3	2	1

No.	Statement	4 always true	3 usually true	2 usually not true	1 never true
17	I am popular in my class	4	3	2	1
18	The temperature in the classroom is comfortable (not to hot not to cold)	4	3	2	1
19	The children in my class make fun of me	4	3	2	1
20	My classroom is well lit	4	3	2	1
21	I am successful in school	4	3	2	1
22	I am satisfied with my marks	4	3	2	1
23	I feel pain or discomfort during the school day (such as in my hands, stomach, back)	4	3	2	1
24	My classroom is quiet enough, so that I can concentrate on my studies	4	3	2	1
25	I would like to transfer to another school	4	3	2	1
26	I like going to school	4	3	2	1
27	My teachers help me succeed	4	3	2	1
28	It is important for me to go to school	4	3	2	1
29	I feel lonely	4	3	2	1
30	My school has a place where it is fun to play (such as a yard)	4	3	2	1
31	School is interesting to me	4	3	2	1
32	There are things in my life that make me frustrated	4	3	2	1
33	I am happy at school	4	3	2	1
34	I have trouble sleeping at night because of the things that happen to me in school	4	3	2	1
35	It bothers me that I don't have things like other children (brand-name clothing, games, equipment, etc.)	4	3	2	1
36	In general, I feel my life is good	4	3	2	1

Appendix A6. Observation checklist.

Formative Assessment Classroom Guide

What part of the lesson are you observing?

First 15 minutes (F)   Middle (M)   Last 15 minutes (L)

Depending on the answer to the above question observers will be looking for different strategies/techniques and therefore only the relevant strategies/techniques should show.

		F	M	L
1	Is it clear what the teacher intends the students to learn? (e.g. sharing learning outcomes with pupils)	/	/	/
2	Does the teacher identify student learning needs? (e.g., eliciting prior knowledge)	/		
3	Do students understand what criteria will make their work successful? (e.g., success criteria, sharing learning objectives, co- constructing success criteria)	/	/	/
4	Are students chosen at random to answer questions?			
5	Does the teacher ask questions that make students think? (e.g., talk partners)	/	/	/
6	Does the teacher give students time to think after asking a question? (e.g., talk partners)	/	/	/
7	Does the teacher allow time for students to elaborate their responses?		/	/
8	Is a whole-class response system used? (e.g., thumbs up strategy)			
9	Is teaching adjusted after gathering pupil feedback (data collection)? (e.g., teacher using immediate pupil feedback to adjust the lesson)		/	/
10	Is there more student talk than teacher talk?		/	/
11	Are most students involved in answering questions?	/	/	/
12	Are students supporting each other's learning? (e.g., talk partners , peer assessment, sharing work under the visualizer)	/	/	/
13	Is there evidence that various forms of teacher feedback advance student learning? (e.g., evidence of pupils editing classwork for improvement, using teacher feedback)	/	/	/
14	Do students take responsibility for their own learning? (e.g. Choosing tasks according to differentiated challenges, Self-assessment, and improvement, planning their own way of working / problem solving , working on or creating new personal targets)	/	/	/

| 15 | Does the teacher give oral formative feedback? | / | / | / |

16 Does the teacher find out what the students have learned before they leave the room? (e.g., talk partners activities, notes from pupils exit messages, questioning)   /

There are 5 possible responses to each of the questions. All questions should be answered. There is only one possible answer to each question.

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented

1. Is it clear what the teacher intends the students to learn?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
When the teacher deliberately defers sharing learning intentions until later	Teacher makes no attempt to communicate learning intentions to students	Teacher states or writes learning intentions but does not check that they make sense to students	Learning intentions clear and communicated to at least half the students	Learning intentions clear and successfully communicated to almost all students

2. Does the teacher identify student learning needs?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Only when the lesson starts with an activity	Evidence of students' prior knowledge would be useful but not elicited	Elicits relevant evidence from only a few students in order to decide where to start the lesson	Refers back to what students know and still need to learn from last period or at start of lesson	Refers back to what students still need to learn and builds on this at the start of the lesson

3. Do students understand what criteria will make their work successful?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
It is always applicable for students to know what makes their work successful	No students appear to know what will make their work successful	Some students are clear about the success criteria for the lesson	Most students are clear about the success criteria for the lesson	Almost all students are clear about the success criteria for the lesson

4. Are students chosen at random to answer questions?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
The teacher asks only whole-class response questions	Students put hands up and teacher chooses one of them to answer a question	Teacher chooses students to answer questions, whether they have their hands up or not	Teacher sometimes uses a randomizing method to choose students to answer questions	Teacher usually uses a randomizing method to choose students to answer questions

5. Does the teacher ask questions that make students think?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Teacher asks no questions	Only recall questions asked	At least one question makes students think	Some questions make students think	Students give thoughtful answers to all questions

6. Does the teacher give students time to think after asking a question?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
When the teacher asks only questions that do not	Teacher passes question to another student, or expands the question if the	Teacher occasionally waits for 3 seconds for a student to think	Teacher usually waits at least 3 seconds before calling on a	Teacher always waits at least 3 seconds before calling on a

| require thinking | student does not answer straight away | before answering a question | student to answer a question | student to answer a question |

7. Does the teacher allow time for students to elaborate their responses?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
When the teacher asks only questions that do not require elaboration	The teacher evaluates or moves on before, or as soon as, student finishes	Teacher occasionally waits at least 3 seconds after student answers a question to allow for elaboration	Teacher usually waits at least 3 seconds after student answers a question to allow for elaboration	Teacher always waits at least 3 seconds after student answers a question to allow for elaboration

8. Is a whole-class response system used?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
This is not a suitable technique for this part of the lesson	Whole class response system not used when it would have been a suitable technique	At least one question asked where all students show their answer, but not all students participate	Some questions asked where all students show their answer, e.g., using fingers, ABCD cards, white boards	Many questions asked where all students show their answer, e.g., using fingers, ABCD cards, white boards

9. Is teaching adjusted after gathering feedback from pupils (data collection)?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
It is appropriate that no data collection is made during this part of the lesson	Teacher makes no adjustments, even though it would have been appropriate	Data collected but not obviously used	Data collected and appeared to be used to adjust teaching to some degree	Teacher comments to class on how data helps her decide on next phase of lesson

10. Is there more student talk than teacher talk?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Only when teacher talk is appropriate for all of this part of the lesson	Teacher talk dominates this part of the lesson, when student talk would have been useful	Students talk as much as the teacher	**Students talk more than the teacher**	Purposeful student talk dominates the time

11. Are most students involved in answering questions?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
No questions asked during the observation	Questions would have been appropriate but not asked	Only a few students involved in answering questions	**About half of the class involved in answering questions**	Most students involved in answering questions

12. Are students supporting each other's learning?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Only quiet individual work observed	It would have been great to see students supporting each other!	Students observed attempting to support other students, but inappropriately	Some students supporting each other effectively	**Many students supporting peers' leaning effectively**

13. Is there evidence that various forms of teacher feedback advance student learning?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Only at start of semester with a new teacher	No teacher feedback	Some teacher feedback, but not helpful	At least one example of feedback used to advance learning	**Many examples of teacher feedback that advances student learning**

14. Do students take responsibility for their own learning?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
It is always applicable for students to be taking responsibility for their own learning	No students are working independently, and no students seek help	Few students are working independently, or seek help at appropriate times	Many students are working independently, or seek help at appropriate times	Almost all students are effectively working independently, or seek help at appropriate times

15. Does the teacher provide oral formative feedback?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Providing oral feedback would not be appropriate	Teacher provides no oral feedback to students despite it being appropriate	Teacher provides oral feedback to students that is unlikely to help them progress	Teacher provides helpful oral feedback to some students	Teacher provides helpful oral feedback to many students

16. Does the teacher find out what the students have learned before they leave the room?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
It is always applicable to find out what students have learned	Students leave room without any feedback to the teacher on what they have learned	Self-reporting by students with no check on the validity of responses	Teacher knows how much some of the students have learned at the end of the lesson	Teacher and students leave the room clear about what has been learned

Appendix A7 Detailed costings of BAU

Appendix A5 Detailed costings of BAU				
Primary Schools	Anglesey	Conwy	Wrexham	
Number of Schools	40	62	67	
Average teacher wage	34034.44	34034.44	34034.44	nasuwt
Salary increase	2%	2%	2%	ASCL
Pension and NI contributions	28%	28%	28%	
Spending on other	16.36%	Does not state take average between the other counties	20.01%	For Wexham they do not have the breakdown
Maintenance fees	9.49%			This is the amount the council calculate that is needed for school (it's part of their funding formula)
Total budget	13,369,000	22,377,000	45,272,030	Does not include SES or ISB
Pupil teacher Ratio (PTR)	18.5	17.7	19.2	
Number of pupils in county	5562	8486	12,343	Council budget document
Number of teachers	199	327	470	Primary only
Average eFSM	15.5	17.5	18	Not needed as comes from different funding area.
	34034.44 28% = 9530	34034.44 28% = 9530	34034.44 28% = 9530	
	34034.44 2% = 681	34034.44 2% = 681	34034.44 2% = 681	
Total salary	44,245	44,245	44,245	
School running costs	25.85	25.85 + 20.01 / 2 = 22.93	20.01	

Total school running costs	13,369,000 25.85 % = 3,360,500	22,377,000 22.93 = 5,044,6000	45,272,030 20.01 % = 9,004,500	
Total	3,360,500	5,044,600	9,004,500	
Total cost per teacher running costs	3,360,500/199 = 16,887	5,044,600 / 327 = 15,426	9,004,500/470 = 19,159	
Total	16,887	15,426	19,159	
Total salary with running costs	44245 + 16,887 = 61134	44245 + 15426 = 59671	44245 + 19,159 = 63,404	
Total	61,134	59,671	63,404	
Including a head teacher in the calculation				
	199 + 40 = 239	327 + 62 = 389	470 + 67 = 537	
	3,360,500/ 239 = 14,061	5,044,600 / 389 = 12,968	9,004500 / 537 = 16,768	
Total salary with running cost and assumption that each school has a head teacher	44245 + 14061 = 58,306	44245 + 12,968 = 56,313	44245 + 16,768 = 61,013	
Total	58,306	56,313	61,013	
Per pupil	58,306/18.5 = 3152	56313/17.7 = 3218	61013/19.2 = 3178	
NASUWT suggest that teacher can only be asked to teach a maximum of 1265 hours over 195 days	58,306/1265 = 46	56313/1265 = 45	61013/1265 = 48	
Cost for a teacher per hour	£46	£45	£48	
Average across all three LAs	**Cost 2018- 2019**	**Cost 2020- 2021**		
Teacher salary	61403	63924		
Teacher salary with HT	58,544	60947		
Cost per pupil (HT inc)	3183	3314		
Cost per hour	46	47.89		

Classroom Climate and Metacognition

CONSIDERATIONS AND RESEARCH

An effective working relationship between teacher and pupil, classroom climate, motivation and metacognition contribute to effective learning
Metacognition and self-regulation mean that pupils think about thinking and know what learning means
Because of this they are able to have more control over their own learning.

Thinking about the way I think, knowing about my own knowledge, understanding my own intelligence, understanding the ways I learn best. Learning about learning.

John Hattie research - effect sizes

Teacher - student relationships   0.72
Metacognitive strategies   0.69

"Too often we teach students what to think, but not how to think" - Dylan Wiliam

"Want to improve your problem solving skills? Try meta-cognition" - OECD Insights

"In the Fixed Mindset when you fail, you're a failure. In the Growth Mindset, when you fail, you're Learning." - Carol Dweck

"We all know that in order for AfL to work, the classroom culture needs to be right as this underpins all the principles of AfL. This is why we began with exploring growth and fixed mindset." - Seamus Gibbons, Headteacher, Langford Primary School

GROWTH MINDSET

- Intelligence can be grown through effort and practice.
- Enjoys difficult work and challenges – it means I'm learning
- Perseverence in the face of challenges
- Learn from criticism/feedback viewed as constructive
- Enjoys learning something new
- Mistakes become a positive thing - learn from them

FIXED MINDSET

- Intelligence is static
- "I have to appear able!"
- Avoids challenges
- Gives up easily
- Looks upon effort as non-beneficial
- Ignores useful criticism
- Reaches a 'plateau' early, underachieves and fails to reach potential

WE ARE ALL A COMBINATION OF THE TWO

www.gwegogledd.cymru

GwE

Talk / Learning / Thinking Partners

Talk partners have been noted to have a strong impact on pupil wellbeing and social skills as well as on peer and self-assessment, pupil to teacher feedback and pupil progress. Effective use of talk partners is an excellent means of gaining knowledge about pupil progress and understanding and to move the learning forward within lessons and over time.

John Hattie research - effect sizes

Assessment literate students	1.44
Providing Formative evaluation	0.9
Classroom discussion	0.82

"If the aim is to teach students self-regulation and control over their own learning then they must move from being students to being teachers of themselves" - John Hattie, Visible Learning for Teachers

Mae Shirley Clarke yn ei nodi fe un o'r 3 ffactor pwysicaf o ran dysgu effeithiol.

"Conditions to be able to function as an active learner – Shirley Clarke
1. A culture of learning, emotional support, climate
   • Growth mindset
   • Metacognition
   • Mixed ability learning
2. Pupils are involved in the planning
3. Talk partners"

"Talk partners create a culture in which pupils do more talking than teachers, cooperatively discussing, answering questions, learning from each other and improving. To maximise their success the following elements need to be in place:
Random pairings which change weekly
Co constructed talk partner success criteria which enable talk training and self and peer evaluation
Random choosing of who answers questions, which eliminates the damaging 'hands up' culture." - Shirley Clarke, Outstanding Formative Assessment - Culture and Practice

"Children need time to think before answering a question. Add discussion time to this and you will see double the impact." - Shirley Clarke, Outstanding Formative Assessment - Culture and Practice

www.gwegogledd.cymru

GwE

216

Involving Pupils in the Planning

Shirley Clarke notes it as one of the 3 most important factors in terms of effective learning.

"Conditions to be able to function as an active learner – Shirley Clarke

1.   *A culture of learning, emotional support, climate*
     *   *Growth mindset*
     *   *Metacognition*
     *   *Mixed ability learning*
3.   *Pupils are involved in the planning*
4.   *Talk partners"*

"The tradition has been that teachers plan lessons away from the children, sometimes asking them what they already know about the subject matter but paying lip service to their involvement. By involving children at the planning stage, their interest and motivation is increased and their achievement is greater because of this ownership." - Shirley Clarke, 2018

Pupils should be given every opportunity to play a part in planning their learning and progress. This involves accommodating pupil voice in planning the content of learning and how they will learn but also in feedback discussions with teachers on their learning progress and next steps.

REASONS FOR INVOLVING PUPILS IN THE PLANNING

*   Pupils are fully active in their learning
*   It keeps the learning visual and interactive
*   It develops pupils' collaborative planning skills
*   Discussions on progress and understanding help the pupil and teacher plan for the next, achievable learning steps
*   It gives pupils a strong voice in their education and a platform for their creativity and imagination
*   It strengthens the relationship between teachers and pupils - teachers respecting their pupils' views and ideas. Interactive teaching can then take place.
*   It reinforces pupils' motivation, independence and confidence when facing new learning.
*   It reinforces effective Formative Assessment principles in the classroom
*   Pupils have ownership of their learning.   Nothing gives a pupil a sense of ownership better than the teacher saying that the lesson/work for the week will explore his/her enquiry/comment/idea.
*   It empowers pupils and develops them as independent learners. It opens the door to creative teaching and enriching pupils' experiences
*   It can involve parents and the wider community in pupils' experiences
*   It creates a breath of fresh air for the teacher. No class is the same, so why use the same lessons/content time after time?

www.gwegogledd.cymru

Differentiated Challenges

Ability grouping, specifically the research and debate against it needs consideration .

"Ability grouping has minimal effects on learning outcomes and profound negative equity effects." [0.12] Hattie, J. 2009

"The evidence is robust and has accumulated over at least 30 years of research ...If schools adopt mixed ability, they are more likely to use inclusive teaching strategies and to promote higher aspirations for their pupils." - Sutton Trust Report 2011

"These studies have repeatedly found that the more schools group by ability, the lower the pupil performance overall." - PISA studies [2012]

"Highly controlled studies suggest that ability grouping has little overall effect on students' academic performance. Greater use should be made of peer co-learning, since these approaches can enhance the learning of all pupils." - Baines, Ed [2012]

"The adoption of structured ability groupings has no positive effects on attainment but has detrimental effect on the social and personal outcomes for some children." - The Primary review, 2008

"Researches in England (Dixon. An Editorial FORUM, 2002) found that 88% of children placed into ability groups at age 4 remain in the same groupings until they leave school. This is one of the most chilling statistics I have ever read." - Joe Boaler 'The Elephant in the Classroom', 2009

Research by John Hattie places this amongst the lowest influences on learning, with an effect size of 0.12

Within class grouping	0.16
Ability grouping	0.12

Mae Shirley Clarke yn ei nodi fel un o'r 3 ffactor pwysicaf o ran dysgu effeithiol:

"Conditions to be able to function as an active learner
1. A culture of learning, emotional support, climate
   - Growth Mindset
   - Metacognition
   - Mixed ability learning

www.gwegogledd.cymru

GwE

Effective Lesson Starters and Questioning to Establish Pupils' Prior Knowledge and Understanding

CONSIDERATIONS AND RESEARCH

The most important feedback in lessons is the feedback that teachers receive from pupils. This is information the teacher receives about pupils' current knowledge and attainment. In order to acquire the correct information, teachers need to start their lessons effectively be means of questioning or starter activities, which will give them this information. There is then a better chance of pupils making progress in their learning during the lesson. Chapter 5 of 'Outstanding Formative Assessment - Culture and Practice' provides useful examples and ideas.

"An effective start to the lesson is a feature of good use of Assessment for Learning. The teacher awakens pupils' interest and draws them in immediately, asking pupils at the start of each lesson how much they know."

"We must plan worthwhile questions that will further deepen pupils' understanding, rather than asking them to recall simple facts" - Shirley Clarke, 2018

"How lessons begin sets the tone for the rest of the lesson – will it be teacher-led or pupils-led, performance-led or learning-led? Will it start with teacher input and quick fired questions, or an immediate task to establish prior knowledge? Will it engage pupils immediately or pave a path of boredom?" - Shirley Clarke, Outstanding Formative Assessment - Culture and Practice

"The most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly" - Ausubel et al 1978

REASONS FOR EFFECTIVE LESSON STARTERS AND ASKING PRIOR KNOWLEDGE QUESTIONS

- The start of the lesson needs to engage pupils and make them think straight away about the content of the lesson
- Finding out what pupils already know or understand, and doing so in every lesson and in every minute of each lesson, is a priority in formative assessment and its main principle. Effective questioning and/or setting a task to ascertain previous learning is therefore essential.
- In so doing, the teacher knows exactly in which direction to go in order to advance pupils' learning, rather than trying to guess this when planning.
- When all pupils undertake such tasks, it gives the teacher a clear picture of each pupils' attainment and the next steps in learning.

QUESTIONING TO ESTABLISH CURRENT KNWOWDGE AND UNDERSTANDING WITHIN AND AT THE END OF LESSONS AR EQUALLY IMPORTANT. THE ABOVE PRINCIPLES APPLY HERE TOO.

www.gwegogledd.cymru

GwE

Using the Visualiser

The visiauliser, or similar classroom tool, is an extremely powerful piece of teaching equipment. It can be used for many formative assessment strategies, e.g. co constructing Success Criteria, modelling excellence, demonstrating a skill, whole class feedback, peer discussion and feedback, self-assessment

John Hattie research - effect sizes:

Assessment literate students	1.44
Providing Formative evaluation	0.9
Classroom discussion	0.82
Feedback	0.72

"Pupils can look at pieces of work completed by their peers. Being able to show a piece of ongoing work directly onto the screen helps pupils in various ways.
1. Modelling excellence as an aid to support others. If pupils see a piece of a high standard covering all the success criteria, pupils are aware of what to aim for.
2. Identification of the elements of the success criteria.
3. Highlighting areas of excellence/ areas to develop
4. As part of a mini-plenary to support and develop pupils' work, whatever the subject area" - Shirley Clarke, 2018

"Mid lesson stops and cooperative marking enable pupils to actively improve their work by seeing excellent examples and dicussing possible improvements"

"This two-step practice discussions leads to pupils working much harder than they used to;" - Shirley Clarke, Outstanding Formative Assessment - Culture and Practice

Three conditions for effective feedback [Sadler]

1. Posess a concept of the goal being aimed for
   [LO and co-constructed SC, more than one excellent example seen]
2. Compare the actual level of performance with the goal
   [Talk partners assess their own work and that of their peers, mid-lesson stops, discuss and improve their own work and borrow ideas]
3. Engage in some appropriate action which leads to some closure of the gap
   [talk partners work together to improve their work after a mid-lesson stop, making amendments straight away]

www.gwegogledd.cymru

GwE

Feedback

CONSIDERATIONS AND RESEARCH

John Hattie research - effect sizes:

Assessment literate students	1.44
Explicit success criteria	0.77
Errors as learning opportunities	0.72
Classroom discussion	0.82
Teacher/ student relationships	0.75
Feedback	0.73
Meta-cognitive strategies	0.60

"The mistake I was making was seeing feedback as something teachers provide to students . I discovered that feedback is more powerful when it is from the student to the teacher ..... Feedback to teachers make learning visible"

and

"Feedback is most effective when students do not have proficiency or mastery – and thus it thrives when there is error or incomplete knowing and understanding. Errors invite opportunity. They should not be seen as embarrassments, signs or failure or something to be avoided. They are exciting, because they indicate a tension between what we now know and what we could know: they are signs of opportunities to learn and they are to be embraced." - John Hattie [2012]

Hattie states that feedback should be:
- *Just for me*
- *Just what I need to help me make progress*
- *Identifies just where I am in my learning process*
- *Just in time.*

The most effective feedback focuses on:
- "The qualities of the child's work and not on comparison with other children
- Specific ways in which the child's work could be improved
- Improvements that the child has made compared to his or her earlier work.'

Crooks [2001]

"Feedback is most effective when goals are specific and challenging but when task complexity is low.
Giving praise for completing a task appears to be ineffective.

www.gwegogledd.cymru

Cognitive Load

CONSIDERATIONS AND RESEARCH

Cognitive Load Theory was developed by John Sweller in 1988
The widely accepted model for Cognitive Load Theory is that the processing of human information is in three parts:

- sensory memory [everything we see, hear and feel]
- working memory
- long-term memory

SENSORY MEMORY

Much of the information that reaches our **sensory memory** [everything we see, hear and feel] is forgotten immediately because we pay little attention to it, or it goes through our working memory and again, most are put to one side only focusing on what is important at the time [*e.g. while working on a writing task, I can hear birds, children playing outside, radio, a door shut, car in the distance, but I'm writing so everything else is put to one side while I'm focusing on that.*]

WORKING MEMORY

It is in the working memory that the handling and the processing of information occurs. That information can come from:

- your sensory memory, or
- as information retrieved from our long-term memory.

The capacity of our working memory is relatively small. We can only process a small amount of information. We can only hold between 5 and 9 items or pieces of information in our working memory at any one time. To be able to think effectively we need enough space in the working memory. We forget information if we don't store it in our long-term memory.

"Understanding of memory, such as an awareness of the limitations of working memory and harnessing the power of long-term memory, is absolutely essential for any educator. Retrieval practice is the process of recalling learned information from long-term memory, with little or no support. Every time a memory is recalled from long-term memory it is altered and strengthened. Recalling information can be challenging and difficult but it is an effective teaching and learning strategy for students of all ages, and is worth embracing" - Kate Jones, 'Retrieval Practice: Primary' [2022]

LONG TERM MEMORY

When we process information, we need to know how to distribute it, then move it to long term memory. The better we can retrieve information from our long-term memory, the more space we have in our working memory to focus on learning new things and solving problems. We can also forget information that is in our long-term memory if we do not retrieve it often enough.

"Memory is the Residue of thought"
To teach well you should pay attention to what assignment will actually make students

www.gwegogledd.cymru

GwE

Learning Outcomes / Intentions

"Pupils can only achieve a learning goal if they understand that goal and can assess what they need to do to reach it." - Sadler 1989

For the whole process of Formative Assessment to be effective, it is essential that the Learning Outcomes are presented to pupils in an appropriate manner, in all activities. The learning outcome should focus on the skill or the knowledge, NOT the activity/task. The LO needs to be out of context and presented in everyday language/childspeak.

DIFFERENT TYPES OF LO
In her recent book '*Unlocking learning Intentions and Success Criteria*' (2021) Shirley Clarke refers to the different types of Learning Objectives / Intentions.

SKILLS [CLOSED]
Decontextualised. Transferable skills. Compulsory elements to be achieved, for example specific mathematical skills, aspects of punctuation. The SC for these usually begin with 'Remember.../You should'
e.g. we are learning to use inverted commas correctly.
Success Criteria: use when a person speaks, titles etc

SKILLS [OPEN]
Decontextualised. Transferable skills. The elements here are more open ended, for example selecting a specific strategy to solve a mathematical problem, conducting a fair test, The SC for these usually begin with 'You could....'
e.g. to create suspense in a piece of writing
The success criteria are a menu/ a toolkit to choose from e.g. write in short sentences, chose effective adjectives

KNOWLEDGE
e.g. we are learning about the history of Tryweryn.
Possible Success Criteria: Know the main events, know the reasons for and against drowning Capel Celyn, impact on local community and residents, on Wales as a whole.
A lesson / series of lessons can include both knowledge and skill LOs.

KNOWLEDGE APPLIED VIA A SKILL
e.g. We are writing a news report (the skill) about the Tryweryn protest in Liverpool (the knowledge)
If we use the above, care needs to be taken when generating the Success Criteria. They will need to be separated for clarity.

www.gwegogledd.cymru

GwE

Success Criteria

CONSIDERATIONS AND RESEARCH

Success Criteria are the breaking down of the Learning Outcome. They provide pupils with steps or ingredients to achieve success. They enable pupils to monitor their own progress and assess their success. Research states that co-constructing SC with pupils is more powerful than a teacher presenting them to pupils.

John Hattie research - effect sizes:

Assessment literate students	1.44
Providing Formative evaluation	0.9
Classroom discussion	0.82
Feedback	0.73

"Once children have success criteria, they have a framework for a formative dialogue, with other peers or teachers, which enable them to:
- *Ensure understanding*
- *Identify success*
- *Determine difficulties*
- *Discuss strategies for improvement*
- *Reflect on progress ."* - Shirley Clarke

FOR 'SKILLS' LEARNING OUTCOMES/ INTENTION, SC MAY BE:
Closed. For a specific skill. E.g. calculate the area of a triangle.
SC for these usually begin with 'Remember to/ You should ...'
or
Open. The elements to be achieved are more open ended. E.g. creating music for a celebration. SC for these usually begin with 'You could...'

FOR KNOWLEDGE LEARNING OUTCOMES/INTENTIONS
These are separate from the decontextualized skills SC. These usually begin with 'You need to know... '

SC do not guarantee quality for open skills (writing, for example); therefore, looking at what excellence looks like is key.
There are now a number of very high-quality techniques not only to get pupils determining success criteria but also to understand what excellence in terms of a specific learning outcome is.

In order to ensure the best possible impact, it is essential that success criteria are:
- co-constructed with pupils, or they will make very little sense and will have less impact on learning
- the same for all learners in the class - differentiation should be ensured by means of how much support is provided in the activity
- used across the curriculum, including work on social and emotional skills

www.gwegogledd.cymru

GwE

NIHR | National Institute for Health Research

PROSPERO
International prospective register of systematic reviews

< Back

Dear Mrs Tiesteel,

Thank you for submitting details of your systematic review "A systematic review of economic evaluations of schools based interventions aimed at improving pupil health, well-being and educational outcomes" to the PROSPERO register. We are pleased to confirm that the record will be published on our website within the next hour.

Your registration number is: CRD42019123564

You are free to update the record at any time, all submitted changes will be displayed as the latest version with previous versions available to public view. Please also give brief details of the key changes in the Revision notes facility. You can log in to PROSPERO and access your records at
https://www.crd.york.ac.uk/PROSPERO

Comments and feedback on your experience of registering with PROSPERO are welcome at: crd-register@york.ac.uk

Best wishes for the successful completion of your review.

Yours sincerely,

PROSPERO Administrator
Centre for Reviews and Dissemination
University of York
York YO10 5DD
t: +44 (0) 1904 321049
e: CRD-register@york.ac.uk
www.york.ac.uk/inst/crd

Appendix B2 Example of terminology

Educational Interventions	Attainment	Pupils	Evaluations	Schools
Programmes	Accomplishment	Children	Measure	Home school
Incentives	Distance Travelled	Students	Analyse	Private school
Improvement programmes	Fulfilment	Young People	Ranking	Public school
Projects	Working towards goals	Minors	Judgment	Faith school
Interventions	Educational Goals	High risk students	Rating	Academies
Improvement projects	Achievement	At risk	Assessment	State run
Improvement	Qualifications	Low achievers	Interpretations	Mainstream
School initiatives		Cognitive outcomes	Impact	International schools
Key Stage 2	Children's Success	Infants	Improvement	High school
Key Stage 3	Level of	Junior	Reporting of interventions	Primary school
Key Stage 4	Completion		Positive effects	Secondary school
	Effects			
	Progress			
Classroom	Summative		Scholastic Attainment	
	Credentials		Children's success	
	Poor readers			
	Closing the Gap			
	Attainment Gap			
	Grades			
	Marks			

Set#	Searched for	Databases	Results
S10	(ab(cost*) OR ab(cost effectiveness*) OR ab(cost benefit*) OR ab(cost utility*) OR ab(cost minimisation*) OR ab(cost analysis*) OR ab(social return on investment*) OR ab(return on investment*) OR ab(Multiple criteria decision analysis) OR ab(economic analysis*)) AND la.exact("English")	Art, Design & Architecture Collection, British Periodicals, Early European Books, GeoRef, Literature Online, Periodicals Archive Online, ProQuest Historical Newspapers: The Guardian and The Observer, PsycINFO , PTSDpubs, SciTech Premium Collection, Social Science Premium Collection, The Guardian, The Vogue Archive	3677527
S12	(ab(education*) OR ab(school*) OR ab(secondary school*) OR ab(primary school*) OR ab(state school*) OR ab(high school*)) AND la.exact("English")	Art, Design & Architecture Collection, British Periodicals, Early European Books, GeoRef, Literature Online, Periodicals Archive Online, ProQuest Historical Newspapers: The Guardian and The Observer, PsycINFO , PTSDpubs, SciTech Premium Collection, Social Science Premium Collection, The Guardian, The Vogue Archive	4055780
S13	(ab(attainment*) OR ab(educational goal*) OR ab(achievement*) OR ab(grade*) OR ab(marks*) OR ab(attainment gap) AND ab(qualification*)) AND la.exact("English")	Art, Design & Architecture Collection, British Periodicals, Early European Books, GeoRef, Literature Online, Periodicals Archive Online, ProQuest Historical Newspapers: The Guardian and The Observer, PsycINFO , PTSDpubs, SciTech	2075111

		Premium Collection, Social Science Premium Collection, The Guardian, The Vogue Archive	
S14	(ab(student*) OR ab(pupil*) OR ab(school child*) OR ab(infant*) OR ab(junior*) OR ab(young people*) OR ab(children*)) AND la.exact("English")	Art, Design & Architecture Collection, British Periodicals, Early European Books, GeoRef, Literature Online, Periodicals Archive Online, ProQuest Historical Newspapers: The Guardian and The Observer, PsycINFO , PTSDpubs, SciTech Premium Collection, Social Science Premium Collection, The Guardian, The Vogue Archive	5592624
S15	((ab(cost*) OR ab(cost effectiveness*) OR ab(cost benefit*) OR ab(cost utility*) OR ab(cost minimisation*) OR ab(cost analysis*) OR ab(social return on investment*) OR ab(return on investment*) OR ab(Multiple criteria decision analysis) OR ab(economic analysis*)) AND la.exact("English")) AND ((ab(intervention*) OR ab(program*) OR ab(school initiatives) OR ab(project) OR ab(study) OR ab(research)) AND la.exact("English")) AND ((ab(education*) OR ab(school*) OR ab(secondary school*) OR ab(primary school*) OR ab(state school*) OR ab(high school*)) AND la.exact("English")) AND ((ab(attainment*) OR ab(educational goal*) OR ab(achievement*) OR ab(grade*) OR ab(marks*) OR ab(attainment gap) AND ab(qualification*)) AND la.exact("English")) AND ((ab(student*) OR ab(pupil*) OR ab(school child*) OR ab(infant*) OR ab(junior*) OR ab(young people*) OR ab(children*)) AND la.exact("English"))	Art, Design & Architecture Collection, British Periodicals, Early European Books, GeoRef, Literature Online, Periodicals Archive Online, ProQuest Historical Newspapers: The Guardian and The Observer, PsycINFO , PTSDpubs, SciTech Premium Collection, Social Science Premium Collection, The Guardian, The Vogue Archive These databases are searched for part of your query.	11073

Yeh, S.S. 20B4 Long version 07 (. The cost-effectiveness of five policies for improving student achievement. American Journal of Evaluation. 28 (4). pp. 416-436.

Checklist	Response
Was a well-defined question posed in an answerable form?	The paper wants to identify of the five interventions under investigations which is the most cost effective at raising pupil's attainment.
Was a comprehensive description of the competing alternatives given?	Two alternatives were described in detail, will full costings for each. Two of the competing alternatives were not described in any detail. There was little description on the normal provision for the control group in the Rapid Assessments description.
Was the effectiveness of the programmes or services identified?	Effectiveness was presented for each alternative. However, given that the effectiveness measure was math and reading the author does not detail what measure were used for the alternatives. The effectiveness of Rapid Assessment was based on a specific program designed test; it is not clear if this is comparable to the other interventions.
Were all the important and relative cost and consequences for each alternative identified?	There is little detail for the costings of the alternatives and the perspective that they took, making it difficult to assess it they were relevant. Some of the interventions e.g., Voucher Programs are aimed at other outcomes not just reading and maths.
Were costs and consequences measures accurately in appropriate physical units?	The author does not detail which measures were used to gain the effect size for the competing interventions, so it is difficult to identify if they were appropriate as an outcome measure. Given the lack of costing detail is difficult to identify if they were accurate or appropriate.
Were costs and consequences valued credibly?	Costs were not measured credibly; teachers time and equipment costs were not included (Only included in sensitivity analysis). It is not clear if the effectiveness measure is a credible measure.
Were costs and consequences adjusted for different timings?	Costs were changed for the 2005 price deflator, discount rates of 3.5% and interest rate of 3% were presented for Rapid assessment but not for the other interventions. There is no follow up data to suggest that Rapid Assessment retained its effectiveness only the assumption that the program would affect would last for 7 years. It is not clear if the other interventions were adjusted.
Was an incremental analysis of costs and consequences of alternative performed?	Only effectiveness cost ratios were presented. There was not incremental cost performed on the cost or effects.
Was allowance made for uncertainty in the establishment of cost and consequences?	Two sensitivity analysis were performed on the Rapid Assessment one with the cost for school to purchase the equipment and one for the cost of teachers to monitor. There was no sensitivity analysis on the effect of the

	intervention, with the author identifying in his notes that of the studies all have different exposure to the intervention. There is no mention of follow up but that the program effects were assumed to last 7 years.
Did the presentation and discussion of study results include all issues of concern?	The author identifies that there are limitations to the conclusion drawn for the analysis of Rapid assessment in terms of the critiques of the role out and the skills that are not measured. There is no presentation of concerns around the different measurements of effect and the reliability of the costings of the other programs.

Yeh S.S. 2009 (A). Shifting the bell curve: The benefits and costs of raising student achievement. Evaluation and program planning. 32 (1). pp. 74-82.	
Checklist	Response
Was a well-defined question posed in an answerable form?	There is not defined question, the author wants to use Benefit Cost Analysis on the intervention of Rapid Assessment. There is no discussion of the method.

Was a comprehensive description of the competing alternatives given?	Given the method used there is no need to have a comparator.
Was the effectiveness of the programmes or services identified?	The effectiveness is based on a previous cost-effective paper. The effectiveness measures are from two studies over a short period. The outcome measure was STAR reading test ad STAR math test, this is a specific test for the intervention.
Were all the important and relative cost and consequences for each alternative identified?	There is good description of the possible long-term benefits of the intervention, extra income, savings to the criminal justice system, welfare savings, and tax savings. Plus, the extra cost to society with the extra collage costs. However not all cost were measured.
Were costs and consequences measures accurately in appropriate physical units?	The measure used to calculate the lifetime earnings was based on AFQT however the measure for Rapid Assessment was program specific, the author makes no attempts to explain the correlation of the measurements through the literature or mapping. So this makes it difficult to see if the measurements were appropriate.
Were costs and consequences valued credibly?	Given the short evaluation on the program 7 month for one study and 18 weeks for the other, the assumptions made from this effectiveness data could be seen as lacking in robustness. The costing of the program is detailed (see previous study Yeh (2007). However, given the issues with S.D and calculation into extra years schooling it is difficult to identify if this is credible.
Were costs and consequences adjusted for different timings?	Yes, all benefits were discounted to age 6, and represent a 44-year working life.
Was an incremental analysis of costs and consequences of alternative performed?	This method does not need this.
Was allowance made for uncertainty in the establishment of cost and consequences?	5 sensitivity analysis were conducted on the effectiveness, the costs (new technology and teacher time), the robustness of the results and the income assumptions.

Did the presentation and discussion of study results include all issues of concern?	The author does not comment or offer recommendation for the reliability of the effectiveness measure, or its comparability to other standardised measures. Given this issue it is not clear that the values are credible.

Belfield, C., Bowden, A.B., Klapp, A., Levin, H., Shand, R. and Zander, S., 2015. The economic value of social and emotional learning. Journal of Benefit-Cost Analysis, 6(3), pp.508-544.	
Checklist	Response
Was a well-defined question posed in an answerable form?	The study is to see if they can use *Benefit Cost Analysis* to mitigate some of the methodological issues in evaluating Social and Emotional learning. This is done by demonstrating correct implementation costings and correct shadow pricing
Was a comprehensive description of the competing alternatives given?	Each intervention was described well along with the limitation of the comparisons.
Was the effectiveness of the programmes or services identified?	Each program had an outcome map detailing all the outcomes both immediate and long-term. Given the difficulties in measuring the outcomes in this area of education the effectiveness is identified for each of the intervention.
Were all the important and relative cost and consequences for each alternative identified?	Costs were established from previous studies, or the authors calculated them from information provided by the authors of the original research there are assumption made on the pricing. There is a good description of where the shadow pricing. Effectiveness has been collated from previous research.
Were costs and consequences measures accurately in appropriate physical units?	Yes, costs were measuring accurately each intervention and had a discussion on the limitations and assumptions of the costs. Costs were inflated to the correct year. Outcomes were converted to monetary values.
Were costs and consequences valued credibly?	Detailed discussion was given to the way in which the authors identified the costing of the outcome into monetary values and were credible.
Were costs and consequences adjusted for different timings?	Yes, program duration and program benefits were adjusted. Discount rates were also applied
Was an incremental analysis of costs and consequences of alternative performed?	This was not needed.

Was allowance made for uncertainty in the establishment of cost and consequences?	The researchers performed 10 sensitivity analysis on the cost and benefits, however not all the programmes had the information to warrant sensitivity in some outcomes.
Did the presentation and discussion of study results include all issues of concern?	The aim was to demonstrate that SEL can be evaluated by the *Benefit Cost Analysis*. This was demonstrated with robust economic methods.

Quinn, B., Van Mondfrans, A.P. & Worthen, B.R. 1984. Cost-effectiveness of two math programs as moderated by pupil SES. Educational Evaluation and Policy Analysis. 6 (1). pp. 39-52.

Checklist	Response
Was a well-defined question posed in an answerable form?	The paper asks two well defined questions: firstly, which pupils benefit most from a particular instructional program? Is it more costly to teach some types of pupils well in one program than another?
Was a comprehensive description of the competing alternatives given?	Both programs were described, there was more detail given to GEMS that Text Maths. It is not clear what Text Math is.
Was the effectiveness of the programmes or services identified?	Each program was evaluated using the same outcome measure, The Iowa test of basic skills (ITBS) and District Maths Test (DMT). Both measures are school wide test and not program specific.
Were all the important and relative cost and consequences for each alternative identified?	There is a comprehensive research design including implantation factors and different levels of implantation. Cost was identified for each intervention.
Were costs and consequences measured accurately in appropriate physical units?	Outcome measures were the same test for each program and used standardised tests (this could essentially be used for other intervention as a comparison) Cost are presented in two tables for the program costs and the implementation costs (they are not aggravated together in tabular form) the author details that the average yearly cost per pupil is $288 for GEMS and $194 for Text Math.
Were costs and consequences valued credibly?	Yes, the research details the cost for each level of SES and Level of Implantation. There is not discount rate only program life there is not discussion of the fall in retention rate of the program
Were costs and consequences adjusted for different timings?	No there was not adjustments made.
Was an incremental analysis of costs and consequences of alternative performed?	No, only cost effectiveness ratios were presented.
Was allowance made for uncertainty in the establishment of cost and consequences?	Sensitivity analysis was not performed on the overall results. The research does discuss sensitivity that could affect the results.

Did the presentation and discussion of study results include all issues of concern?	Yes, the researchers discussed that GEMS was more cost effective for higher rate SES and that Text Math was more cost effective for the lower SES. They were able to identify the issues of data collection and the limitations that the data they collected had. There suggests area for improvement for further analysis of the GEMS maths program. The author does identify in the discussion that they could have implemented outcome measure for pupil ability and teacher moral, as well as better data on previous attainment. But does not address there being no sensitivity analysis.

Reynolds, A., Temple, J., Robertson, D. & Mann, E. 2002. Age 21 cost-benefit analysis of the Title I Chicago Child-Parent Centres. Educational Evaluation and Policy Analysis. 24 (4). pp. 267-303.	
Checklist	Response
Was a well-defined question posed in an answerable form?	They are testing the hypothesis if they can calculate economic benefits at age 21 for the participants in the Title I Chicago Child-Parent Centres.
Was a comprehensive description of the competing alternatives given?	Given that this was a cost benefit there is no need to detail alternatives. They use data from the control group to compare the outcomes. There was a control group, there was not a detailed discussion of the treatment as usual
Was the effectiveness of the programmes or services identified?	The main outcome measure was high school graduation.
Were all the important and relative cost and consequences for each alternative identified?	Yes, the cost and consequences were clear and appropriately identified for the research aims. The cost for normal provision was not reported.
Were costs and consequences measures accurately in appropriate physical units?	There is a wide range of shadow prices used for the lifetime benefits of the intervention and are appropriate. Saving to remedial services, lifetime earnings, tax revenues for federal and state. Reductions in the cost of crime within the criminal justice system. There was shadow pricing for the savings to adult victims of crime and child abuse averted. The costs were measured appropriately and were based on previous research.
Were costs and consequences valued credibly?	Yes, the cost and consequences were clear and appropriately identified for the research aims. The cost for normal provision was not reported
Were costs and consequences adjusted for different timings?	Discount rates were applied at 3% this is a follow up study of longitudinal data so there is no need to adjust for timing.
Was an incremental analysis of costs and consequences of alternative performed?	No this is not needed for this type of economic evaluation.

Was allowance made for uncertainty in the establishment of cost and consequences?	Sensitivity analysis was performed and presented for different ages of the pupils in relation to the discount rate. There was no other sensitivity analysis conducted.
Did the presentation and discussion of study results include all issues of concern?	The authors conclude that the program is beneficial to the participants and the wider society. There discuss the results in line with other research that has been conducted on other pre-school programs. There is a comprehensive discussion on the limitation and the policy implication of the results.

Bowden, A.B. and Belfield, C., 2015. Evaluating the Talent Search TRIO program: A Benefit-Cost Analysis and Cost-Effectiveness Analysis 1. Journal of Benefit-Cost Analysis, 6(3), pp.572-602.

Checklist	Response
Was a well-defined question posed in an answerable form?	They want to identify if the Talent Search program is cost effective across all sites and is there a cost benefit to the program. This is the first economic evaluation of this program.
Was a comprehensive description of the competing alternatives given?	The comparison for the cost effectiveness element of the research, did not give a good description of the costs associated with the intervention across the sites.
Was the effectiveness of the programmes or services identified?	Each site had the effectiveness measured in both high school graduation and post-secondary outcomes. The main effectiveness data comes from previous research by (Constantine, Seftor, Martin, Silva & Myers, 2006)
Were all the important and relative cost and consequences for each alternative identified?	All the costs and consequences were measured for each site. There is not cost for business as usual other than suggest that the business as usual would not cost more that talent search.
Were costs and consequences measures accurately in appropriate physical units?	Given the purpose of the program the cost and measurements were appropriate for the research. For the outcomes measure for the cost benefit analysis. They use the life-time model of high school failure from Karoly (2015) earning profiles were from Current Population Survey and were calculated on the gender differences and age.
Were costs and consequences valued credibly?	The valuation of the cost and consequences are creatable however in the note (11) they did not count the cost of the office of post-secondary education who oversee the running of the project. They also did not count the cost to train the Talent Search staff. Given that this is essential to the running the program the author did not explain counting.
Were costs and consequences adjusted for different timings?	The effects and costs were discounted at 3%. The adjusted for the multi-year program for up to 7 years and discounting at 3%. Recourses that lasted longer than a year, such as facilities and computers, were amortized using a rate of 5%

Was an incremental analysis of costs and consequences of alternative performed?	No there were none performed on the results.
Was allowance made for uncertainty in the establishment of cost and consequences?	Comprehensive sensitivity tests were conducted for each economic method, on both the cost and the effects.
Did the presentation and discussion of study results include all issues of concern?	The authors discuss the implementation issues and why there could be large disparities in a) the costs and b) the outcomes. They also see that there is little information on what treatment as usual is for pupils not receiving Talent Search.

Hollands, F., Bowden, A.B., Belfield, C., Levin, H.M., Cheng, H., Shand, R., Pan, Y. and Hanisch-Cerda, B., 2014. Cost-effectiveness analysis in practice: Interventions to improve high school completion. Educational Evaluation and Policy Analysis, 36(3), pp.307-326.

Checklist	Response
Was a well-defined question posed in an answerable form?	This was an example of how cost effectiveness can be used in education. They authors wanted to highlight the importance of this for policy making and how is can aid decision makers.
Was a comprehensive description of the competing alternatives given?	Each intervention was described so that the reader understood what each intervention entailed
Was the effectiveness of the programmes or services identified?	Effectiveness and cost were sought from different research articles and personal correspondence with program designers.
Were all the important and relative cost and consequences for each alternative identified?	The costs were measured from a societal perspective. Reasonable adjustments were made in the costing of the intervention where specific data was not available or national rather that local prices needed to be calculated.
Were costs and consequences measures accurately in appropriate physical units?	To make the interventions comparable the author used the HSD or GED as the main outcome indicator however they do suggest that there is a difference in these indicators as a HSD is more desirable for future employers.
Were costs and consequences valued credibly?	The Talent Search intervention worked with students who were on track to complete their education in school anyway. The other interventions were for students that had dropped out of school already. All the interventions are targeted at reducing drop out, but the Talent Search has an advantage over the other interventions as the students had not dropped out.
Were costs and consequences adjusted for different timings?	Costs were adjusted to 2010 prices. A discount rate was not applied to the other interventions as they lasted a year and the outcome was measured at the end of the year. Tales Search was discounted at 3% and as the average time of the program was 1-7 years.
Was an incremental analysis of costs and consequences of alternative performed?	Costs were presented in two ways the first was a classic cost effectiveness ration per point effect for dollars. The other ration was per $100,000 x amount of HSD or GED.
Was allowance made for uncertainty in the establishment of cost and consequences?	Sensitivity analysis was presented for the Talent Search program but not the others. They do discuss the uncertainty in the costings of the interventions as it was retrospective.

Did the presentation and discussion of study results include all issues of concern?	The authors discuss that the cost and outcome for some interventions are large, and the cost effectiveness analysis is not able to compare all the benefits of the intervention. They detail that there are inconsistencies in collecting cost data that target different population in different locations. That address the issues of comparability of Talent Search as the target population are still in school. They also suggest that Cost Benefit analysis would be appropriate to the other interventions.

Levin, H.M., Glass, G.V. and Meister, G.R., 1987. Cost-effectiveness of computer-assisted instruction. Evaluation review, 11(1), pp.50-72.	
Checklist	Response
Was a well-defined question posed in an answerable form?	The research sought to answer the question is Computer-Assisted Instruction cost effective that other intervention that are targeting attainment of students
Was a comprehensive description of the competing alternatives given?	All interventions were described in good detail.
Was the effectiveness of the programmes or services identified?	Effectiveness data was collated from other research for all of the other interventions.
Were all the important and relative cost and consequences for each alternative identified?	Costs and consequences were reported from other evaluations. The cost and effects from CIA are reported from a meta-analysis of 14 studies that the author calculated (Levin, 1984). The costs can be seen tabulated in (Levin 1984) in detail. For the Cross Age Tutoring the author retrospectively costed the program, and they were estimates. For reducing class size there were also estimates in that they assuming that maths and reading would take two thirds of the teaching time so divided the cost by three to identify a cost per subject. The issue with this is that they could have been effects in other subject or additional benefits to a smaller class size, but this was not taken into consideration. The authors calculate the estimates for each reduction in class numbers using a formula however they do not detail the subsets from the Glass and Smith, 1979 meta-analysis that they use.
Were costs and consequences measures accurately in appropriate physical units?	The author did not present the units in which the effect sizes were measured. There is no detail if they were standardised tests or program specific tests. All cost were measured retrospectively and by the authors, they detail where some of the costs have come from but not all.

Were costs and consequences valued credibly?	The costs were valued credibly to answer the research question, however given the lack of detail in how the effect size was measured if make it difficult to identify if the tests were appropriate to answer the research question and comparable.
Were costs and consequences adjusted for different timings?	The author does not give any detail in costing for different times of discount rates for each intervention. There is a discussion about cost could be reduced further as the price of computed may fall but this is not a robust calculation.
Was an incremental analysis of costs and consequences of alternative performed?	None were calculated for this research.
Was allowance made for uncertainty in the establishment of cost and consequences?	None were reported.
Did the presentation and discussion of study results include all issues of concern?	The author does advise caution when interpreting the results of the research, this is that the result is of a general analysis of each program. They suggest that other analysis of the program (using different data) could have had other results. They suggest that the costs and effects could be different in different school and with different populations

Hollands, F.M., Kieffer, M.J., Shand, R., Pan, Y., Cheng, H. & Levin, H.M. 2016. Cost-Effectiveness Analysis of Early Reading Programs: A Demonstration with Recommendations for Future Research. Journal of Research on Educational Effectiveness. 9 (1). pp. 30-53	
Checklist	Response
Was a well-defined question posed in an answerable form?	The research aim is to see if cost –effectiveness analysis can be applied to early reading interventions.
Was a comprehensive description of the competing alternatives given?	The two interventions were from the same research project and they both were random assignment to each intervention. Description of the interventions were given.
Was the effectiveness of the programmes or services identified?	Yes, as both interventions were being evaluated at the same time that the same outcome measure for the interventions. They looked at alphabetic, text reading and comprehension.
Were all the important and relative cost and consequences for each alternative identified?	Yes, the costs and consequences are identified for each intervention. Details costing for each intervention are presented.
Were costs and consequences measures accurately in appropriate physical units?	The measurement tool is the same for each intervention and they measure appropriately reading levels. Costs were measured appropriately.
Were costs and consequences valued credibly?	Yes, the cost and consequences are measures credibly giving a breakdown of what witch element on the intervention costs and the percentage that the cost was to the whole intervention.
Were costs and consequences adjusted for different timings?	Costs were converted to 2010 prices and the discount rate of 3%for cost incurred over 2–3-year lifetime of the program. Costs were sought from intervention staff and national prices available for e.g., square food of school building development.
Was an incremental analysis of costs and consequences of alternative performed?	None were presented for this research. Only Cost effectiveness ration for each intervention.

Was allowance made for uncertainty in the establishment of cost and consequences?	Sensitivity analysis was performed on the length of time the students were exposed to the intervention as the developers recommended. No sensitivity analysis was presented for the costing of the interventions.
Did the presentation and discussion of study results include all issues of concern?	The authors discuss that there could be issues if the teachers were newly qualified in comparison to experience teachers. They discuss the methodological challenges in using cost-effectiveness in evaluation interventions and reducing the effectiveness to one outcome. They also recommend that cost analysis is built into the evaluation framework so that the costing of intervention can be more accurate. They also suggest that the research community and developers, policy makers and so on see the importance of cost effectiveness in research on educational interventions. So, this can aid decision makers in the take up of interventions.

Yeh, S.S. 2009. Class size reduction or rapid formative assessment? A comparison of cost-effectiveness. Educational Research Review. 4 (1). pp. 7-15.	
Checklist	Response
Was a well-defined question posed in an answerable form?	The aim of this paper is to evaluate if Class Size Reduction is cost effective compared to Rapid Assessment.
Was a comprehensive description of the competing alternatives given?	The author gave a good description of each intervention.
Was the effectiveness of the programmes or services identified?	The effectiveness of the intervention was from Greenwald et al.'s (1996) for the class size reduction. The author also gave effect sizes for other evaluation and a comprehensive discussion of why they effect sizes where discounted from the analysis. The rapid assessment results were from (Nunnery, Ross, & McDonald, 2006; Ross, Nunnery, & Goldfeder, 2004 for the effect size for reading and Ysseldyke & Bolt, 2007 and Ysseldyke & Tardrew, 2007 for the effects size for maths. the mean for each was calculated
Were all the important and relative cost and consequences for each alternative identified?	Cost for the Class size reduction were calculated from Reichardt (2000) and cost for Rapid Assessment were from Yeh (2007). Reading and maths outcome were the only consequences that were measured. Given that reducing class size could have impacted other areas of student achievement this was not explored.
Were costs and consequences measures accurately in appropriate physical units?	Standardised tests were used to measure the effectiveness of Class Size Reduction, however the Rapid Assessment used program specific effectiveness measure. It is not clear if these are comparable as effectiveness measures. There is not a detailed list of cost and where the cost fell.
Were costs and consequences valued credibly?	The costs were not valued in any detail and with no breakdown of the costs it is difficult to estimate if the costs were credible for either intervention. The effectiveness data from Rapid Assessment was conducted over a short period (less than a year) there is no follow up or longitudinal data to support the effects lasting over time
Were costs and consequences adjusted for different timings?	Costs were adjusted to 2006 and inflation adjustment were made using the 1997 price index. There is no adjustment made for the effect size, but the author does discuss different effects from different studies.

Was an incremental analysis of costs and consequences of alternative performed?	None were presented, Relative Effectiveness cost ratios were presented.
Was allowance made for uncertainty in the establishment of cost and consequences?	The author also details a relative *effectiveness- cost ratio.* The *effectiveness-cost ratio* for Rapid Assessment is 124 times as large as the effectiveness-cost ratio for class size reduction. In other words, student achievement would increase 124 times faster for every dollar (or euro) invested in rapid assessment rather than Class Size Reduction. No sensitivity analysis was performed on the costs or the effects.
Did the presentation and discussion of study results include all issues of concern?	The author did discuss the robustness of the evaluation and weather like Class Size Reduction, Rapid Assessment could have had the Hawthorn effect as has been suggested in the effect sizes of some evaluations of Class Size Reduction. The author does not address the relatively short evaluation time of Rapid Assessment or the issues that could be with the testing of the effect size. The author does not discuss the measurements issues.

Borman, G.D. and Hewes, G.M., 2002. The long-term effects and cost-effectiveness of Success for All. Educational Evaluation and policy analysis, 24(4), pp.243-266.

Checklist	Response
Was a well-defined question posed in an answerable form?	The research wanted to address the hypothesis that the long-term effects would be evident and that the cost would support the financial argument for prevention rather that remediation. They wanted to demonstrate that Ramey and Ramey's (1998) six principal framework that the effects of the intervention would be evident in the retention of the effect size in later schooling.
Was a comprehensive description of the competing alternatives given?	The analysis was completed on the original intervention and control schools. There is not a description of treatment as usual for the control school but the author did detail that the control schools were in less need of intervention as they were not ranked as ' in need of improvement' . The second analysis of comparing the intervention a description was given.
Was the effectiveness of the programmes or services identified?	Effectiveness of the programmes were measured in math and reading results on standardised tests.
Were all the important and relative cost and consequences for each alternative identified?	The author presented the cost in a tabulated from and gave clear description of where the costs were obtained. The author explains that the costs were discounted for the other intervention as they targeted other outcomes and did not just target maths and reading. The costing of the other interventions was obtained from previous research.
Were costs and consequences measures accurately in appropriate physical units?	Yes, reading and maths outcome were measured using appropriate measures. The cost of pupil who avoided special educational rescores where costed in detail.
Were costs and consequences valued credibly?	Yes, all costs were valued credibly for a study of this type of the author detailed the savings to special educational provision that was credible. The effects were measured credibly also.
Were costs and consequences adjusted for different timings?	Costs were converted to 2000 dollars price; the other intervention were analysed at a reduced cost to reflect the comparators programs effect on reading and math (personal correspondence with the program developers)
Was an incremental analysis of costs and consequences of alternative performed?	None were presented for this research.

Was allowance made for uncertainty in the establishment of cost and consequences?	None were conducted on the costs and consequences.
Did the presentation and discussion of study results include all issues of concern?	The author discussed school wide intervention are not solely focused on maths and reading. They discuss contextual differences based on location and staff recruitment for further implantation. Also discussed was that some of the interventions (especially pre-school) have not been able to fully see the benefits later on. They addressed the difference in design as this was quasi experiments and the other interventions were experimental. They suggest that cost benefit analysis would have been appropriate for this intervention given that the benefits were long term.

Bowden, A. B., Belfield, C. R., Levin, H. M., Shand, R., Wang, A., & Morales, M. (2015). A benefit-cost analysis of City Connects. Centre for Benefit-Cost Studies of Education, Teachers College, Columbia University.	
Checklist	Response
Was a well-defined question posed in an answerable form?	The questions were well defined and clear reflected what the research wanted to achieve.
Was a comprehensive description of the competing alternatives given?	As this was Benefit cost the comparator was not detailed.
Was the effectiveness of the programmes or services identified?	The effectiveness was based on a previous study, and this was then modelled to calculate a yield of high school graduates
Were all the important and relative cost and consequences for each alternative identified?	All relative cost and consequences were detailed, the author identified benefits outside of the remit of the research question giving an in-depth look at the consequences of the intervention. The costing of the intervention was very detailed and done on three different models to identify where each cost and benefit would be seen.
Were costs and consequences measures accurately in appropriate physical units?	Cost and consequences were measured in appropriate terms.
Were costs and consequences valued credibly?	The cost and consequences were valued to reflect to effect and the true cost to all parties involves in the intervention, costs were derived from many different resources in a systematic and credible way.

Were costs and consequences adjusted for different timings?	Cost and consequences were adjusted at 3.5 % over a 6-year period.
Was an incremental analysis of costs and consequences of alternative performed?	None were presented.
Was allowance made for uncertainty in the establishment of cost and consequences?	Sensitivity analysis was performed on the costs, and a discussion of conservative estimated for effects were also presented.
Did the presentation and discussion of study results include all issues of concern?	Author identifies that there could be limitations to the research that they conducted, and the sample of community partners was small (10%) They discuss the difficulties of shadow pricing with this type of research. They suggest that further scale up of the program is recommended. That to further the investigation there should be wider benefit measures other that education attainment and the variability of the cost for each site.

Appendix B5 PRISMA preferred reporting checklist

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	

Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	

255

DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

For more information, visit: **www.prisma-statement.org**.

C1 Consent for interviews and focus groups.

COLEG GWYDDORAU IECHYD AC YMDDYGIAD
COLLEGE OF HEALTH & BEHAVIOURAL SCIENCES

YSGOL SEICOLEG
SCHOOL OF PSYCHOLOGY

Dear Class Teacher,

Re: Evaluating the impact of GwE Formative Assessment Project.
Information sheet for Teachers

Bangor University is currently undertaking an evaluation of the impact of formative assessment approaches as part of the GwE Shirley Clarke project. Your school has be chosen to take part in the research and that one or more of the class teachers will receive formative assessment training. We would like to invite you to participate in a short interview to help us further understand the benefits of this important project.

Previous research has identified that formative assessment is successful at raising attainment of pupils. However, there is lack of research detailing the wider impact of the program on pupils, the teacher and the classroom culture as a whole. We would like to conduct a short interview with you to help us gather information about the impact of this project.

Please find enclosed an information sheet that provides more information about the research, why it is being carried out, and what it will involve. There is also a consent form attached for you to complete if you wish to participate in this project.

If you still have any questions or concerns please do not hesitate to contact the researchers involved in the project (contact details below).

Yours faithfully

Emma Tiesteel

Emma Tiesteel	PhD Student	hbpc14@bangor.ac.uk	07810838393
Prof Carl Hughes	Project Supervisor	c.hughes@bangor.ac.uk	01248 388275
Dr Joanna Charles	Project Supervisor	j.charles@bangor.ac.uk	01248 388896
Dr Richard Watkins	Project Supervisor	richardwatkins@gwegogledd.cymru	0300 500 8087

COLEG GWYDDORAU IECHYD AC YMDDYGIAD
COLLEGE OF HEALTH & BEHAVIOURAL SCIENCES

YSGOL SEICOLEG
SCHOOL OF PSYCHOLOGY

Title of the study: *Evaluating the Impact of Formative Assessment on Primary School Pupils*

Project team:
Emma Tiesteel (PhD student, School of Health Care Sciences, Bangor University)
Prof Carl Hughes (Project supervisor, School of Psychology, Bangor University)
Dr Joanna Charles (Project supervisor, CHEME, Bangor University)
Dr Richard Watkins (Project supervisor, GwE)

Information about the study
Teachers within your school will receive training on the effective use of formative assessment strategies. Bangor University intends to evaluate the impact of this project on the pupils and the standards they achieve. Previous research has indicated that this strategy has been successful at raising the attainment of pupils, however there is little known research detailing the wider impact that formative assessment has on pupils.

This research is seeking consent for you to be monitored over two years to look at the wider impact of this training. We would like to ask you to take part in a short one-to-one interview with the researcher about your experiences in the classroom. This will take place at the end of the 2018-2019 academic year, and again as a follow-up interview during 2019-20 academic year.

Why have I been contacted?
As your school will receive training provided by GwE, and your school has indicated that they want to take part in the research being conducted by Bangor University, we are contacting you to see if you would like to take part in the research.

Are there any benefits or risks?
We do not believe that yourself or the children in you class will be put at risk in any way. Should there be any signs of distress the questionnaires or short interview will be stopped. The benefits to taking part in the research will be that your work within the classroom will add to the knowledge of formative assessment.

What will happen to my data?

All data collected will be confidential, you and the school will not be identified in any report, thesis, or publication that may arise from this study. Data from this study will contribute towards research on evaluating the impact of formative assessment. If you choose to withdraw from the study we will ignore information collected from yourself and not use it in any way. Data will be destroyed seven years after the project is finished this is in line with new data protection regulations.

What if I don't want to take part?

It is your decision to participate in this study. There will be no consequences should you wish to not take part. You can withdraw your data from the study at any time without giving a reason.

Who do I contact about the study?

If you have any further questions, please contact Emma Tiesteel (the PhD student conducting the research), or the project supervisors: Prof Carl Hughes, Dr Joanna Charles or Dr Richard Watkins. Names are contact details are listed below:

Emma Tiesteel	PhD Student	Hbpc14@bangor.ac.uk	07810838393
Prof Carl Hughes	Project Supervisor	c.hughes@bangor.ac.uk	01248 388275
Dr Joanna Charles	Project Supervisor	j.charles@bangor.ac.uk	01248 388896
Dr Richard Watkins	Project Supervisor	richardwatkins@gwegogledd.cymru	0300 500 8087

Who do I contact with any concerns about this study?

If you have any concerns or complaints about this study, or of the conduct of the individuals conducting this study, then please contact Mr Huw Ellis, School Manager (huw.ellis@bangor.ac.uk)

RESEARCH CONSENT FORM FOR TEACHERS

Please return this sheet if you would like to take part in the research outlined.

Please initial box

I wish to take part in a short one to one interview

I understand that participation is voluntary

I understand that I can withdraw my data at any time

Please sign and return this form,

Name in block letters:...

Signature:... Date.................

If you have any further questions, please contact the supervisor of the project, Dr Carl Hughes (Tel: 01248 383278; email: c.hughes@bangor.ac.uk) , Dr Joanna Charles (01248 388896 ; email : mhs002@bangor.ac.uk) or Dr Richard Watkins (Tel: 0300 500 8087; email: richardwatkins@gwegogledd.cymru).
If you have any complaints about how this study is conducted please address these to: Mr Huw Ellis, School Manager (huw.ellis@bangor.ac.uk)

Dear·Parent/Guardian,¶

·Re:·Evaluating·the·impact·of·Shirley·Clarke·Formative·Assessment·Project.·
Information·sheet·for·Parents/Guardians

Bangor·University·is·currently·working·with·schools·in·North·Wales·to·evaluate·the·quality·of·feedback·pupils·receive·in·school.·This·project·is·being·undertaken·in·partnership·with·the·Regional·School·Effectiveness·and·Improvement·Service·for·North·Wales·(GwE).·Your·child's·school·has·been·chosen·to·be·part·of·the·project.·We·would·now·request·consent·for·your·son/daughter·to·help·us·further·understand·the·benefits·of·this·important·project.·

Previous·research·has·identified·that·improving·feedback·from·teachers·improves·the·quality·of·learning·experiences·in·classrooms.·However,·there·is·lack·of·research·detailing·the·wider·impact·of·the·program·on·pupils,·the·teachers·and·the·associated·benefits·to·the·region's·schools.·

Please·find·enclosed·an·information·sheet·that·provides·details·about·the·research,·why·it·is·being·carried·out,·and·what·it·will·involve.·There·is·also·a·consent·form·attached·for·you·to·complete·if·you·wish·your·child·to·participate·in·this·project.·If·you·still·have·any·questions·or·concerns·please·do·not·hesitate·to·contact·the·researchers·involved·in·the·project·(contact·details·below).¶

Yours·faithfully¶

Emma·Tiesteel·¶

Emma·Tiesteel	PhD·Student	Hbpc14@bangor.ac.uk	07810838393
Prof·Carl·Hughes	Project·Supervisor	c.hughes@bangor.ac.uk	01248·388275
Dr·Joanna·Charles	Project·Supervisor	j.charles@bangor.ac.uk	01248·388896
Dr·Richard·Watkins	Project·Supervisor	richardwatkins@gwegogledd.cymru	0300·500·8087

¶
COLEG·GWYDDORAU·IECHYD·AC·YMDDYGIAD¶
COLLEGE·OF·HEALTH·&·BEHAVIOURAL·SCIENCES¶
¶
YSGOL·SEICOLEG¶
SCHOOL·OF·PSYCHOLOGY¶‖
¶

Project·team·¶

Emma·Tiesteel··(PhD·student,·School·of·Health·Care·Sciences,·Bangor·University)¶
Prof··Carl·Hughes·(Project·supervisor,·School·of·Psychology,·Bangor·University)·¶
Dr·Joanna·Charles·(·Project·supervisor,·CHEME,·Bangor·university)¶
Dr·Richard·Watkins·(Project·supervisor,·GwE)·¶
¶

Information·about·the·study¶

Teachers·within·your·child's·school·will·receive·training·to·understand·how·to·improve·the·
quality·of·feedback·they·provide·in·the·class.·Bangor·University·are·working·with·GwE·to·
evaluate·the·impact·of·this·project.··Previous·research·has·indicated·that·these·approaches·have·
been·successful·at·raising·the·attainment·of·pupils,·however·there·is·little·known·research·
detailing·the·wider·impact·of·the·programme·on·pupils,·teachers·and·the·associated·benefits·to·
the·region's·schools.·¶
¶

This·research·is·seeking·consent·for·your·child·to·be·monitored·over·two·years·to·look·at·the·
wider·impact·of·this·training.·We·would·like·to·ask·pupils·to·take·short·questionnaires·at·the·
beginning·and·end·of·the·project,·together·with·small·group·discussions·with·a·researcher·about·
their·experiences·in·the·classroom.··¶
¶

Why·have·I·been·contacted?¶

Your·school·will·receive·training·provided·by·GwE,·and·the·school·has·indicated·that·they·want·
to·take·part·in·the·research·being·conducted·by·Bangor·University.·We·are·contacting·you·to·see·
if·you·would·like·your·child·to·take·part·in·the·research.¶
¶
¶

Are·there·any·benefits·or·risks?¶

We·do·not·anticipate·any·risks·associated·with·participation.·However,·should·there·be·any·signs·
of·distress·during·the·questionnaires·or·group·discussion·the·activity·will·be·stopped.·In·this·case·
an·appropriate·member·of·school·staff·will·be·informed·and·your·child·will·be·cared·for·in·the·
appropriate·manner.·¶
¶
¶

263

What·will·happen·to·my·child's·data?¶

All·data·collected·will·be·confidential,·your·child·will·not·be·identifiable·in·any·report,·thesis,·or·
publication·that·may·arise·from·this·study.·Data·from·this·study·will·contribute·towards·research·
on·evaluating·the·impact·of·formative·assessment.·If·you·choose·to·withdraw·your·child's·data·
from·study·we·will·remove·and·safely·dispose·of·any·information·collected·on·your·child·and·

What if I don't want my child to take part?

It is your decision to allow your child to participate in this study. There will be no consequences should you wish for your child to not take part. You can withdraw your child's data from the study at any time without giving a reason.

Who do I contact about the study?

If you have any further questions, please contact Emma Tiesteel (the PhD student conducting the research), or the project supervisors: Prof Carl Hughes, Dr Joanna Charles or Dr Richard Watkins. Names and contact details are listed below:

Emma Tiesteel	PhD Student	Hbpc14@bangor.ac.uk	07810838393	
Prof Carl Hughes	Project Supervisor	c.hughes@bangor.ac.uk	01248 388275	
Dr Joanna Charles	Project Supervisor	j.charles@bangor.ac.uk	01248 388896	
Dr Richard Watkins	Project Supervisor	richardwatkins@gwegogledd.cymru	0300 500 8087	

Who do I contact with any concerns about this study?

If you have any concerns or complaints about this study, or of the conduct of the individuals conducting this study, then please contact Mr Huw Ellis, School Manager (huw.ellis@bangor.ac.uk)

·······················Page Break·······················¶

265

COLEG GWYDDORAU IECHYD AC YMDDYGIAD
COLLEGE OF HEALTH & BEHAVIOURAL SCIENCES

YSGOL SEICOLEG
SCHOOL OF PSYCHOLOGY

RESEARCH CONSENT FORM FOR PARENTS/GAURDIANS

Please return this sheet if you ==would== like your child to take part in the research outlined. This will not affect yours or your child's legal right or education.

	Please initial box	
I agree for my child to take part in the questionnaires		
I agree for my child to take part in the small discussion groups		
I agree for my child's achievement data to be collected		
I understand that participation is voluntary		
I understand that I can withdraw my child's data at any time		

Please sign and return this form to your child's school at your earliest convenience.

Child's Name in block letters:...¶

Please indicate what school year group your child is in:..................................¶

Parent / Guardian Signature:...Date.................¶

Parent / Guardian Name in block letters:...¶

If you have any further questions, please contact the supervisor of the project, Dr Carl Hughes (Tel: 01248 383278; email: c.hughes@bangor.ac.uk), Dr Joanna Charles (01248 388896; email: j.charles@bangor.ac.uk) or Dr Richard Watkins (Tel: 0300 500 8087; email: richardwatkins@gwegogledd.cymru).¶

Appendix C2 Topic Guide

Topic guide

Teacher Interviews
Topic guide.
Pre interview:

- Explain that the interview is confidential and that they will not be identified in any way. So they can be completely honest.
- Explain that they can stop at any time, should they want to.
- Ask if they have any questions,
- Have they signed the consent form?
- We are only focusing on what is new in the school.

Did you know about SC before?

What elements did you focus on or were new to the classroom practice?

- Growth mind-set: Do the Pupils understand and can explain this. Using the muscles to forward there learning.
- Learning outcomes: Out of context learning should be always seen
- Success criteria: Co constructed and visible to the class. Examples
- Talk partners: Random selection, weekly change.
- Differentiated challenges/ mixed ability grouping: For all pupils, for all lessons
- Visualizer: Pit stops, uses examples of work, peer feedback
- Successful starts to lessons, prior knowledge questions.
- Involving pupils in the planning: Independence within their own learning
- Elimination of rewards: barriers or facilitators: Praise effort, Celebrate mistakes
- Feedback: Have they change the way in which they give feedback, has it reduced the workload

Effect in the class/school both academic and behavioural Social Effects

Personal/professional development and attitudes

Challenges to this way of working

Dissemination in the school

What do you think of the peer mentoring model of delivery?

How much time did it take you to implements as in extra time on top of your normal prevision? Did it save time e.g. more effective marking.

What extra materials were needed to deliver the SC?

Is it now just part of the way that you work?

Have you had any feedback from parents?

Appendix DI Control Consent

Dear·Parent/Guardian,

Re:·Researching·the·impact·of·the·GwE·Formative·Assessment·Project.
Information·sheet·for·Parents/Guardians

Bangor·University·is·currently·working·with·schools·in·North·Wales·to·evaluate·the·quality· of·feedback·pupils·receive·in·school.·This·project·is·being·undertaken·in·partnership·with· the·Regional·School·Effectiveness·and·Improvement·Service·for·North·Wales·(GwE).·Your· child's·school·has·been·chosen·to·be·part·of·the·research.·We·would·now·request·consent· for·your·son/daughter·to·complete·a·short·survey·to·help·us·further·understand·the· benefits·of·this·important·program.·

Previous·research·has·identified·that·improving·feedback·from·teachers·improves·the· quality·of·learning·experiences·in·classrooms.·However,·there·is·lack·of·research·detailing· the·wider·impact·of·the·program·on·pupils,·the·teachers·and·the·associated·benefits·to·the· region's·schools.·

Please·find·enclosed·an·information·sheet·that·provides·details·about·the·research,·why·it· is·being·carried·out,·and·what·it·will·involve.·There·is·also·a·consent·form·attached·for·you· to·complete·if·you·wish·your·child·to·participate·in·this·project.·If·you·still·have·any· questions·or·concerns·please·do·not·hesitate·to·contact·the·researchers·involved·in·the· project·(contact·details·below).

Yours·faithfully

Emma·Tiesteel·

Emma·Tiesteel	PhD·Student	Hbpc14@bangor.ac.uk	07810838393	
Prof·Carl·Hughes	Project·Supervisor	c.hughes@bangor.ac.uk	01248·388275	
Dr·Joanna·Charles	Project·Supervisor	j.charles@bangor.ac.uk	01248·388896	
Dr·Richard·Watkins	Project·Supervisor	richardwatkins@gwegogledd.cymru	0300·500·8087	

271

Title of the study: *Evaluating the Impact of Formative Assessment on Primary School Pupils*

Project team

Emma Tiesteel (PhD student, School of Health Care Sciences, Bangor University)
Prof Carl Hughes (Project supervisor, School of Psychology, Bangor University)
Dr Joanna Charles (Project supervisor, CHEME, Bangor University)
Dr Richard Watkins (Project supervisor, GwE)

Information about the study

Teachers within your child's school are waiting to take part in training to understand how to improve the quality of feedback they provide in the class. Bangor University are working with GwE to evaluate the impact of this project. Previous research has indicated that these approaches have been successful at raising the attainment of pupils, however there is little known research detailing the wider impact of the programme on pupils, teachers and the associated benefits to the region's schools.

This research is seeking consent for your child to be monitored over two years to look at the wider impact of this training. We would like to ask pupils to take short questionnaires at the beginning and end of the project, together small group discussions with a researcher about their experiences in the classroom.

Why have I been contacted?

Your child's school has indicated that they wish to take part in the research being conducted by Bangor University, we are contacting you to see if you would like your child to take part in the research.

Are there any benefits or risks?

We do not anticipate any risks associated with participation. However, should there be any signs of distress during the questionnaires or group discussion the activity will be stopped. In this case an appropriate member of school staff will be informed and your child will be cared for in the appropriate manner.

What will happen to my child's data?

All data collected will be confidential, your child will not be identifiable in any report, thesis, or publication that may arise from this study. Data from this study will contribute towards research on evaluating the impact of formative assessment. If you choose to withdraw your child's data from the study we will remove and safely dispose of any information collected on your child and not use it in any way. Your child's data will be securely stored and destroyed seven years after the project has finished this is in line with new data protection regulations.

273

It is your decision to allow your child to participate in this study. There will be no consequences should you wish for your child not to take part. You can withdraw your child's data from the study at any time without giving a reason.

Who do I contact about the study?

If you have any further questions, please contact Emma Tiesteel (the PhD student conducting the research), or the project supervisors: Prof Carl Hughes, Dr Joanna Charles or Dr Richard Watkins. Names are contact details are listed below:

Emma Tiesteel	PhD Student	Hbpc14@bangor.ac.uk	07810838393
Prof Carl Hughes	Project Supervisor	c.hughes@bangor.ac.uk	01248 388275
Dr Joanna Charles	Project Supervisor	j.charles@bangor.ac.uk	01248 388896
Dr Richard Watkins	Project Supervisor	richardwatkins@gwegogledd.cymru	0300 500 8087

Who do I contact with any concerns about this study?

If you have any concerns or complaints about this study, or of the conduct of the individuals conducting this study, then please contact Mr Huw Ellis, School Manager (huw.ellis@bangor.ac.uk)

·······················Page Break·····················

275

RESEARCH CONSENT FORM FOR PARENTS/GAURDIANS

Please return this sheet if you would like your child to take part in the research outlined. This will not affect yours or your child's legal right or education.

	Please initial box
I agree for my child to take part in the questionnaire's	
I agree for my child to take part in the small discussion groups	
I agree for my child's achievement data to be collected	
I understand that participation is voluntary	
I understand that I can withdraw my child's data at any time	

Please sign and return this form to your child's school at your earliest convenience.

Child's Name in block letters:..

Please indicate what school year group your child is in:..............................

Parent / Guardian Signature:...Date..................

Parent / Guardian Name in block letters:..

Intervention Pro forma for School Budgeting Framework/Resource Expenditure Framework

Name of intervention	
Main focus: *e.g., reading, mathematics, behaviour*	
Description of intervention *Provide a descriptive narrative about the intervention, including a practical description of how it is used in school.*	Description: Data collection: Approach: (Universal, Targeted, Individual)

What is the aim of the intervention? *Describe what the intervention/programme is designed to improve, e.g., standards of reading, spelling or procedural numeracy skills.*	
Mechanisms of delivery: *Describe delivery method, including the time allocation and setting.*	Who delivers: How often: Where: How many pupils benefit from the intervention (across one academic year):

Evidence of effectiveness	Academic research, Programme website, Toolkit (EEF), Consortia, in school data or other?
What is the overall cost?	*e.g. TA cost and/or newly purchased resources or accessories, and supply cover cost if required.* *Training:* *Resources: (Printing, books, room hire…)* Cost to purchase or annual licence renewal: *Other: (Supply cover, transport…)*
Does this intervention meet the aims? *In this section schools are encouraged to reflect on the*	**What is the evidence** In School data:

evidence for impact of the intervention/programme. It might be useful to think of your evidence in a hierarchy: - Discussions about social validity, i.e., did teachers and pupils value the intervention/programme? Do they view it positively? - Survey responses from teachers/pupils - Pupil progress data.	From teachers: Form teaching assistants: From pupils:	
Do we implement this intervention/programme/strategy well?	*The intervention is implemented as recommended*	🟩
	The Intervention has been adapted somewhat	🟨
	The intervention has been largely adapted to meet the needs of the school	⬛
	The intervention is not implemented as recommended	🟥
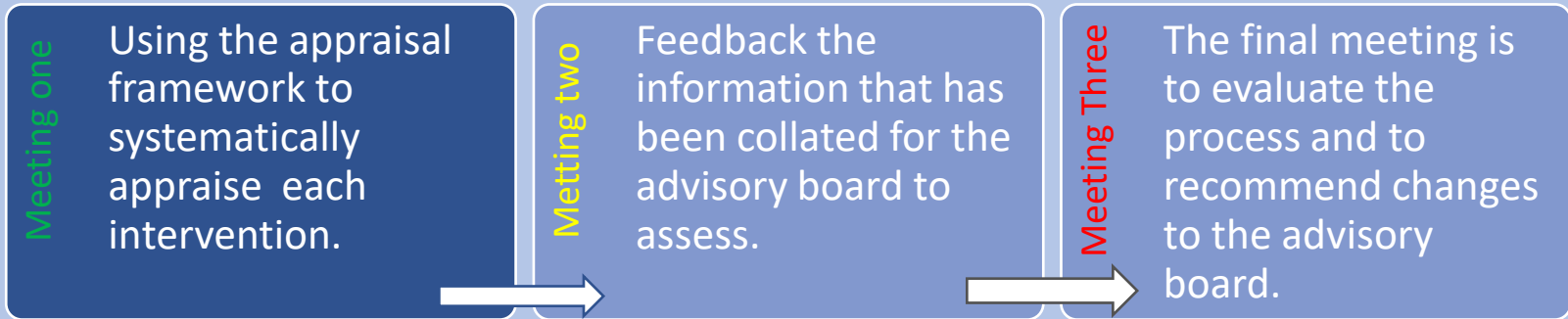
Does the intervention improve the attainment of disadvantaged pupils?		
	The initiative improves the attainment of eFSM pupils	🟩
	There is some evidence that this initiative improves the attainment of eFSM pupils	🟨
	There is very limited evidence that this initiative improves the attainment of eFSM pupils	⬛
	There is no evidence that this initiative improves the attainment of eFSM pupils and none of the features of this intervention are likely to close the attainment gap	🟥
Additional Information		

Rhyl Cluster Programme Budgeting Marginal Analysis (PBMA)

A priority appraisal framework

Emma Tiesteel
PhD researcher Bangor University hbpc14@bangor.ac.uk

School	
Introduction and aims of the project.	The aim of this project is to support schools in prioritising pupils' outcomes in a robust and evidenced informed procedure. This will be done using the methodology of Programme Budgeting Marginal analysis. This is done by identifying an advisory board within the school consisting of staff who are responsible for decision making for financial or the employment of interventions to target pupils' outcomes. The process of PBMA is spread over three meetings with the advisory board to support decision making on the provision within the school.

Meeting one Using the appraisal framework to systematically appraise each intervention. → **Metting two** Feedback the information that has been collated for the advisory board to assess. → **Meeting Three** The final meeting is to evaluate the process and to recommend changes to the advisory board.

Each intervention that is employed with in the school will be assessed on the evidence of effectiveness based on the iWa-RLC project, evidence of social validity, implementation and costs to run the intervention are all components of the appraisal framework. The information is then imputed into a matrix and categorised. The school then receive a report to digest and discuss, following this the board will vote on how to procced with each intervention. This final stage is to make recommendations and evaluate the process. The advisory board will make decisions for each intervention from the option below:
A.   Keep current program
B.   Expand current program
C.   Limit current program
D.   Seek an alternative |
| **What is PBMA?** | Programme Budget and Marginal Analysis (PBMA) is a systematic decision-making process used in healthcare science to identify cost-effective interventions. |

PBMA is used in the NHS to help decision makers maximise patients' outcomes with the needs of the local population, within the specified budget. This is done by apprising resource allocation and outcomes of specific programmes. Marginal analysis consists of apprising the added benefits and the costs if new programmes are invested in, taking into consideration the loss of benefits if other programmes are disinvested.
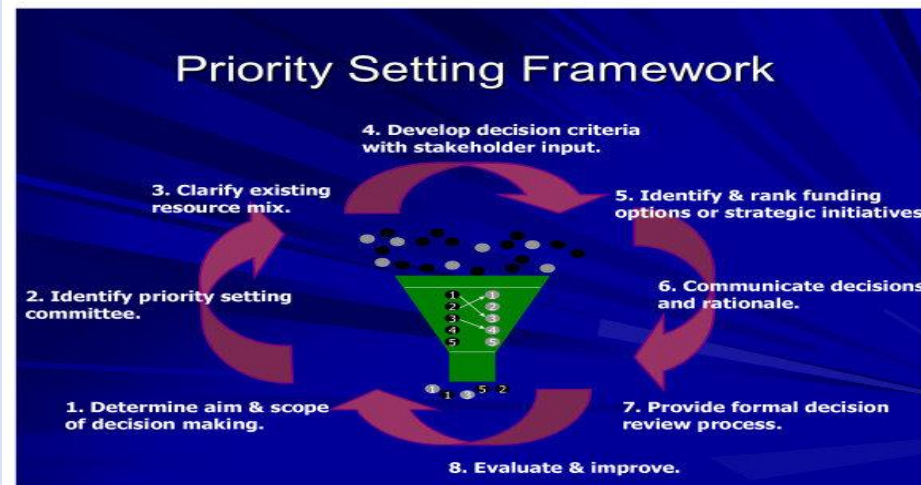
The aim of this project is to appraise a variety of academic and well-being interventions that are being delivered in school and use PBMA to help school leaders improve their understanding of the cost and effectiveness of the interventions/programmes. To achieve this aim, a wide range of evidence will be evaluated and synthesised, including evidence of effectiveness, evidence of costs, quality of implementation, evidence of impact and evidence of social validity. This framework has not been used in the primary or secondary education sector but is used extensively in health-care settings. The longer-term aim of this study is to enable school leaders and policy makers to use the skills from this project to inform future decision making, including the use of increasingly scarce financial resources.

Who are the Advisory Board members?	Deputy head teacher Additional Learning Needs (ALN) Coordinator Teaching Assistant (Grade 5)

Which interventions are being evaluated during this study?	1. Say All Fast Minute Every Day Shuffled -SAFMEDS 2. Indirect Dyslexia Learning- IDL 3. Social and Communication group 4. Narrative Therapy 5. ALN Literacy Support	6. Nurture Group 7. Lunchtime club 8. Social and Emotional afternoon group 9. Emotional Literacy Support Assistant - ELSA
How do we assess evidence of effectiveness? (Information was collated from the iWaB-RLC project) See Appendix 1 for evidence of effectiveness categorisation. See Appendix 2 for methodology for calculating costs	3* 2* NL2	NA Ex - Excluded from iWaB-RLC report NI- Not on iWaB-RLC report
How do we evaluate costs?	Detailed cost for all interventions/programmes are contained in Appendix 3.	General resources and staff costs are contained in Appendix 4.
How do we evaluate the aims of the intervention? Does it have an impact and how is it received (social validity)	🟩 There is evidence of effectivness in school - staff and pupils have a positive attitude towards the intervention/programme 🟨 Evidence of staff and pupil enjoyment 🟧 No data to support evidence of effectivness 🟥 Staff or Pupils do not enjoy the intervention	

How do we implement the intervention ?	<table><tr><td>🟩</td><td>The intervention is implemented as recommended</td></tr><tr><td>🟨</td><td>The Intervention has been adapted somewhat</td></tr><tr><td>🟧</td><td>The intervention has been largely adapted to meet the needs of the school</td></tr><tr><td>🟥</td><td>The intervention is not implemented as recommended</td></tr></table>
Table of recommendations	Based on the school's individual responses to the interventions that were candidates for appraisal, recommendations are made to maximise the outcomes for the pupils that receive the intervention.

Main focus	Name of intervention	Description of intervention	What is the aim of the intervention	Mechanisms of delivery	Evidence of effectiveness	Cost per pupil	Does the intervention meet the aims	Do we implement the intervention well	What percentage of pupils receive eFSM for this intervention
Mathematics	SAFMEDS	Intervention to improve maths fluency	Improved standards of numeracy /math fluency	3 groups of pupils for 3 mornings a week (6 pupils in each group) 20 mins prep per week	3*on iWa-RLC project	**£73.87**	Although results have been variable over recent years, last year's results showed an improvement. TA's and pupils enjoy the intervention	Intervention is implemented as recommended	28%
Reading and Spelling	Indirect Dyslexia Learning (IDL)	Computer based individualised learning at pupil's level	Improve reading and spelling	One session for 15 minutes with a group of up to 15 students. 2 TA's monitor the session with 70 pupils in total having the intervention. Each child works individually at a laptop.	NI	£68.32	Increase in reading / spelling standardised scores/RA/SA. Pupils progress data. Pupils seem content with the intervention. TAs sometimes face challenges delivering the intervention due to time constraints.	Intervention is implemented as recommended	40%

Main focus	Name of intervention	Description of intervention	What is the aim of the intervention	Mechanisms of delivery	Evidence of effectiveness	Cost per pupil	Does the intervention meet the aims	Do we implement the intervention well	What percentage of pupils receive eFSM for this intervention
Social and Communication Skills	Social and Communication group	An alternative provision for ASD/ social communication pupils. Covering basic skills in Literacy and number in an autism friendly environment	Improve social/ communication skills which will enable pupils to access the curriculum	2 hours daily with 3 TAs. One TA is grade 3 and the other two are grade 2.	2* on iWa-RLC project	£1247.25	Pupils enjoy it. Parents/carers are all positive. TA's enjoy delivering it. Boxall (pre and post) Children are able to cope better in a larger mainstream environment.	School designed intervention	22%
Speech and Language	Narrative therapy	Small group provision 2x 1 hr for KS2 (8 pupils) and 1 x for FPh (6 pupils)	Improve speech and Language (Narrative) skills	3 1hr sessions per week Grade 5 TA per session	2* on iWa-RLC project	£103.23	Staff and pupils enjoy the sessions. Unsure of impact on learning. Evaluation using pre- and post-intervention data not currently available	Intervention is implemented as recommended	64%

Main focus	Name of intervention	Description of intervention	What is the aim of the intervention	Mechanisms of delivery	Evidence of effectiveness	Cost per pupil	Does the intervention meet the aims	Do we implement the intervention well	What percentage of pupils receive eFSM for this intervention
Reading /Spelling/Phonics	ALN Literacy support in small (including Teaching Handwriting Reading and Spelling Skills - THRASS)	Reading/Spelling literacy support in small groups with TA	Improve S/S in reading or spelling	Foundation phase receive 4 hrs a week with 23 pupils in total. KS2 receive 6 hrs per week with 21 pupils in total.	NI	£99.52	Data is collected on Pupils progress and they enjoy the intervention. Pupils and TAs enjoy the intervention.	The intervention has somewhat been adapted (Teaching Handwriting Reading and Spelling Skills - THRASS)	49%

Main focus	Name of intervention	Description of intervention	What is the aim of the intervention	Mechanisms of delivery	Evidence of effectiveness	Cost per pupil	Does the intervention meet the aims	Do we implement the intervention well	What percentage of pupils receive eFSM for this intervention
Well-being	Nurture Group	Well-being intervention aimed at pupils who need time to settle into school in the mornings.	To support vulnerable pupils to transition into school	Two groups: one Foundation Phase (5 times a week for 30 mins with a grade 5 TA and 8 pupils) the other in KS2 (4 sessions a week for 30 mins with a grade 5 TA and 10 pupils)	2* on iWa-RLC project. Used PASS for identification. Boxall profile.	£302.77	Pupils enjoy intervention and BOXALL evidence. Improvements noted on pupils' scores.	The Intervention has been adapted somewhat	44%
Well-being, to accommodate children who struggle with school skills and behaviour on the playground	Lunchtime club	10 children per day with room for a drop in for children who are not coping that day	To keep the child safe and happy in a friendly relaxed environment	It is run from 12.00-1.00, but some children just need to play or eat for half an hour	Excluded from on iWa-RLC project	£281.20	Children definitely benefit from the small provision; behaviour improves, and pupils are ready to settle into afternoon sessions	The Intervention has been adapted somewhat	29%
Main focus	Name of intervention	Description of intervention	What is the aim of the intervention	Mechanisms of delivery	Evidence of effectiveness	Cost per pupil	Does the intervention meet the aims	Do we implement the intervention well	What percentage of pupils receive eFSM for this intervention

Behaviour and Social/Emotional	Social and emotional afternoon group.	Small group of 6 pupils for 5 afternoons a week (2hrs)	To improve social and emotional skills and behaviour.	Delivered in the main building by a grade 5 TA 10 hrs a week in total.	Not on iWa-RLC project report	£791.38	Pre/Post questionnaires Boxall/PASS/Teacher used to evaluate. Data from previous year indicated most children made expected/good/very good progress. Pupils enjoy the sessions and the TA really enjoys delivering the intervention. Parental feedback is good.	School designed intervention with no recommended implementation.	50%
Well-being	ELSA	Delivered for two full afternoons a week for7-10 weeks. There are 5 children on the course.	To improve pupils' well-being, confidence, self-esteem, friendship and self-regulation skills. Also helps address anger management and problems at home and in school	Run in a small teaching room in a quiet, relaxed setting. Each child has up to one hour	NL2 on iWa-RLC project	£106.08	The TA enjoys delivering the sessions and pupils give very positive feedback. We do a pre/post form which shows how much improvement (if any) has been made.	Implemented as recommended	83%

Recommendations

Results :	In total 3 academic interventions were candidates for appraisal • Math :(n-1) • Spelling/phonics/reading: (n-2) In total 2 Additional Learning Needs (ALN)interventions were candidates for appraisal • Social and Communication group • ALN Literacy In total 4 wellbeing interventions were candidates for appraisal: • Nurture Groups • Lunchtime club • Social and emotional afternoon group • Emotional Literacy Support Assistant - ELSA
Costings range:	Academic interventions cost per pupil range from £68.32 - £103.23 Wellbeing interventions cost per pupil range from £106.08 - £791.38 ALN interventions cost per pupils range from £99.52- £1247.25
Costing range across the Rhyl cluster:	Academic interventions cost per pupil range from: £37.27- £302.59 Wellbeing interventions cost per pupils range from: £28.05- £791.38 ALN interventions cost per pupils range from: £99.52- £1247.25

293

Areas of concern:	Although results have been variable over recent years, last year's results showed an improvement. Teaching assistants enjoy delivering the intervention and pupils enjoy the intervention
SAFMEDS	
Recommendations:	School data needs to be scrutinised to gather the effectiveness on the academic outcomes of the pupils that receive the intervention.

Areas of concern:	Positive in-school data to support the impact on the pupil's progress. There are issues with the time constraints for the delivery of the intervention.
Indirect Dyslexia Learning (IDL)	
Recommendations:	For the impact to continue adequate resources need to be addressed.

Areas of concern: **Narrative Therapy**	**Staff enjoy the delivery of the intervention and pupils enjoy the sessions. There is currently no data available to support impact. (Pre and post evaluations are carried out)**
Recommendations:	**School data needs to be scrutinised to gather the effectiveness of the academic outcomes of the pupils that receive the intervention.**
Areas of concern: **Lunchtime club**	**There is anecdotal evidence that the pupils benefit from the provision.**
Recommendations:	**The school would be advised to collect robust data to support the provision of the Lunchtime club**

Appendix 1	**Evidence of effectiveness categorization**: * indicates intervention has evidence of impact. Evidence grades in are based on the criteria used by the EIF (Early Intervention Foundation), which uses five levels of strength of evidence to evaluate the degree to which a programme has been shown to have a positive, causal impact on specific child outcomes. These are: • Level 4 recognizes programmes with evidence of a long-term positive impact through multiple rigorous evaluations. At least one of these studies must have evidence of improving a child outcome lasting a year or longer. • Level 3 recognizes programmes with evidence of a short-term positive impact from at least one rigorous evaluation, that is, where a judgment about causality can be made. • Level 2 recognizes programmes with preliminary evidence of improving a child outcome, but where an assumption of causal impact cannot be drawn. • Not Level 2 (Not Level 2) distinguishes programmes whose most robust evaluation evidence does not meet the level 2 threshold for a child outcome. • Not Effective (found not to be effective in at least one rigorously conducted study). This is reserved for programmes where a rigorous programme evaluation (equivalent to a level 3) has found no evidence of improving child outcomes or providing significant benefits to other participants. This rating should not be interpreted to mean that the programme will never work, but it does suggest that the programme will need to adapt and improve its model, learning from the evaluation. • No Evidence Available (NA), no evidence currently available.
Appendix 2	**SAFMEDS standardised costs**: It was necessary to standardise the cost of SAFMEDS as there were large discrepancies for each school. Expert information was sought from Kaydee Owen a research with extensive knowledge of the intervention (Owen, et al., 2020, under review). Each pupil is expected to progress through 10 packs of card in an academic year. Card 50 £0.94 Printing 50 £0.15 — £1.99 per pack X 10 = £19.90 per set per pupil = Cost per year £6.63* Laminating 50 £0.90 *The cost was annuitized under the assumption that each set of cards will last 3 years. Each school is different in the approach to time to print and make the packs. Preparation time is calculated based on the information that was provided to the researcher.

Appendix 3

Detailed cost breakdown for each intervention:

SAFMEDS *			
Number of students 3 groups of 6 pupils	18		
	Unit	Cost	Total per year
TA time 2.25 hours per week	2.25	£21.06	£821.34
Prep Time 20 mins per week	0.33	£3.12	£121.68
Training *	0	0	0
Resources *		£221.40	£73.80
Contribution to member of staff in the cluster *			£312.89
Number of pupils			
Cost per year			**£ 1329.71**
Cost per pupil			**£73.87**
* Calculations are for the intervention being delivered in a group setting. 15 mins per group 3 times a week for 3 groups.			
* No training cost were reported for this intervention			
*: Resources were calculated at 10 packs per pupil at £19.90 per year. The cost is divided into three on the assumption that the pack will last three years.			
*: SAFMEDS TA was calculated one day a week at 9.36 per hour times by 39 weeks then divided by the 7 schools			

Indirect Dyslexia Learning (IDL)			
Number of students, two groups of 35 students	70		
	Unit	Total	Total per year
TA time (Grade 2) * for 10 hours a week	10	£93.60	£3,650.40
Prep time 20 min per day total 1.66 per week	1.66	£15.53	£605.96
Assessment time Per year	35		£327.60
Licence fee			£199
Training *	0	0	0
Cost per year			£4,782.96
Cost per pupil			£68.32
*: No training cost were reported for this intervention.			
*: TA was costed at Grade 2, no details were given. Cost for grade 5 would be £82.05 per pupil			

Social and Communication Group			
Number of students: Small group intervention	9		
	Unit	Total	Cost per year
TA time per week* - 2 hours a day with 3 TAs	30	£282.70	£11,025.30
Resources	1	£200	£200
Training *	0	0	0
Cost per year			**£11,225.30**
Cost per pupil			**£1,247.25**
*: Cost include two Grade 2 TAs (£9.36 per hour) and one Grade 3 (£9.55 per hour)			
*: This is a school designed intervention			

Narrative Therapy			
Number of pupils, one group has 6 pupils, one group has 8 pupils.	14		
	Unit	Cost	Cost per year
TA Time (Grade 5) KS2 has a 2-hour session per week and foundation phase has 1 hour per week	3	£33.57	£1,309.23
Prep time	0.25	£2.79	£109.10
Resources*		£140	£28
Training *	0	0	0
Cost per year			£1,445.33
Cost per pupil			£103.23
*: Programme specific resources were divided by 5 under the assumption that they would last 5 years before more investment is needed.			
*: No training cost were reported			

ALN Literacy Support			
Number of pupils, 23 in foundation phase and 21 in KS2		44	
	Unit	Cost	Cost per year
TA Time (Grade 5)- 4 hours a week for foundation phase, and 6 hours a week for KS2 pupils.	10	£111.90	£4,364.10
Paper	1	£1.89	£1.89
Printing	1	£1.50	£1.50
Books *	1	£11.70	£11.70
Training *	0	£0.00	0
Cost per year			**£4,379.19**
Cost per pupil			**£99.52**
*: Cost of books were from :https://www.englishphonicschart.com/s-97-english-spelling-chart-pack-training-pack-for-teachers-assistants-students-and-parents-197-p.asp			
*: No training cost were reported			

	Unit	Total	Total per year
Social and Emotional afternoon group.			
Number of Pupils, small group intervention			6
TA Time (Grade 5) 5 afternoons a week ,2 hour per day.	10	£111.90	£4,364.10
Resources (Books)	1	£100	£100
Ream of paper	1	£1.89	£1.89
Prep time,30 mins pre time a week	0.5	£5.59	£218.20
Assessment time 30 min per pupil twice a year	6	£67.14	£67.14
Training *	0	0	0
Cost per year			**£4,748**
Cost per pupil			**£791.38**
*: School designed intervention no training cost were reported.			

Emotional Literacy Support Assistant - ELSA			
Number of Pupils, 1:1 intervention	15		
	Unit	Total	Total per year
TA (Grade 5) 1 hour a week on the average of 8.5 sessions.	42	£569.98	£1,409.94
Prep time 20 mins per week 2.8 hours over the average 8.5 sessions	2.8	£31.33	£93.99
Assessment time 20 mins per pupil	1.6	£17.90	£53.70
Training*	0	£0.00	0
Supervision	3	£33.57	£33.57
Resources	0	£0.00	0
Number of sessions	8.5		
Cost per year			£1,591.20
Cost per pupil			£106.08
*: Each pupil has between 7-10 sessions 1:1. This is averaged out at 8.5 sessions for each child. The assumption that this intervention is run 3 time a year with 5 children each term having the intervention			
*: No training cost were reported (Not from school budget)			

Nurture group			
Number of pupils, Foundation phase 5 pupils, KS2 5 pupils.	10		
	Unit	Total	Yearly total
TA time, 30 mins daily for foundation phase and 30 4 times a week for KS2	4.5	£50.35	£1,963.84
Prep time, 30 mins per week	0.5	£5.59	£218.20
Assessment time, 30 mins per pupils per year.	9	£100.71	£100.71
Resources (Food)	1	£10	£390
Training*	1	£1,775	£355
Total cost per year			**£3027.75**
Cost per pupil			**£302.77**
*: Staff training is divided over 5 years under the assumption that due to staff turnover re training will be needed. The cost of cover, travel and accommodation are not included in the calculation.			

Lunch time club			
	Unit	Total	Total per year
TA time (Grade 5), one hour daily.	5	£55,95	£2,182.05
Resources *	0	0	0
Number of pupils	10		
Cost per year			£2,182.05
cost per pupil			£218.20
*No resources were recorded.			

Appendix 4

General resources	
Staffing Costs correct as of January 2020.	TA hourly rate: £9.36 HLTA hourly rate: £12.35 Teacher hourly rate: £24 - £32 per hour, depending on salary point TA supply cover (agency costs): £80 per day Teacher supply cover: £170 per day
Resources Costs correct as of January 2020.	Printing/photocopying: the schools are charged roughly 0.003p per unit (so £1.50 for 500 pages) for toner etc. Copier Paper: £1.89 per ream (it's currently £9.46 for a box of 5 reams) Laminating pouches: £1.81 for a box of 100 Pens: £0.80 for a box of 50 White card is £1.88 for a pack of 100, or £3.99 for a pack of 200
Training Costs	With consultation of the Rhyl cluster heads training costs were annuitized at 5 years. This is under the assumption that extra resources would be needed due to staff turnover or retraining.

THANK YOU TO THE SCHOOL AND ALL THE STAFF WHO WERE ABLE TO MAKE THIS REPORT POSSIBLE

For further information please contact Emma Tiesteel: hbpc14@bangor.ac.uk

Indexing

1 Barriers to implementation

    1.2 Resistance from staff
    1.3 Resistance from pupils
    1.4 Resistance from SLT
    1.5 Resistance from parents
    1.6 School ethos (resistance to change)
    1.7 School context
    1.8 Classroom makeup
    1.9 Resistance from teacher
    1.10 Other

2 Facilitators to implementation :

    2.1 Behaviour change (reflexive practice)
    2.2 SLT support
    2.3 Parental involvement
    2.4 Other staff involvement
    2.5 Improving strategies already implemented
    2.6 Pupils involvement
    2.7 Other

3 Action research

    3.1 Sharing ideas with other schools and teachers (f)
    3.2 Not fostering a working relationship
    3.3 Logistics
    3.4 Language
    3.5 Cluster working

4 Benefits to strategies

    4.1 Independent learning
    4.2 Focused pupil own learning
    4.3 Understanding pupils learning (minute by minute)

Emma Tiesteel
First category was deleted

Emma Tiesteel
this category was emarginated into the above.

Emma Tiesteel
This was deleted and moved to school context

Emma Tiesteel
this was a category that was after deeper inspection.

Emma Tiesteel
This was changed to other strategies

Emma Tiesteel
Deleted only one comment moved to another section

Emma Tiesteel
Deleted other school getting more support. Merged the firs column.

Emma Tiesteel
Language and cluster working were moved to logistics.

Emma Tiesteel
Move data to confidence column. Self esteem and self

Appendix E

Test of normality for quantitative data

Test of Normality CHU-9D.

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Condition	Statistic	df	Sig.	Statistic	df	Sig.
Utility	Intervention	.122	188	.000	.882	188	.000
	Control	.118	220	.000	.914	220	.000
a. Lilliefors Significance Correction							

Test of Normality for SDQ

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Condition	Statistic	df	Sig.	Statistic	df	Sig.
Total difficulties score	intervention	.123	170	.000	.958	170	.000
	control	.124	184	.000	.958	184	.000

a. Lilliefors Significance Correction

Test of normality QoLS

		Kolmogorov-smirnov[a]			Shapiro-wilk		
Condition		Statistic	Df	Sig.	Statistic	Df	Sig.
Qols_StuTeach	Intervention	.205	138	.000	.805	138	.000
	Control	.202	140	.000	.783	140	.000
Qols_Phys	Intervention	.106	138	.001	.944	138	.000
	Control	.088	140	.010	.969	140	.003
Qols_SocPsych	Intervention	.106	138	.001	.959	138	.000
	Control	.129	140	.000	.931	140	.000
Qols_PosAtt	Intervention	.103	138	.001	.940	138	.000
	Control	.139	140	.000	.934	140	.000

A. Lilliefors significance correction

All ANOVA results for QoLS.

Tests of Between-Subjects Effects

Dependent Variable: QoLS_StuTeach

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.814[a]	3	.271	1.626	.184
Intercept	3724.109	1	3724.109	22304.684	.000
CONDITION	.638	1	.638	3.822	.052
TIME	.059	1	.059	.352	.554
CONDITION * TIME	.119	1	.119	.711	.400
Error	45.749	274	.167		
Total	3770.163	278			
Corrected Total	46.563	277			

a. R Squared = .017 (Adjusted R Squared = .007)

b. Computed using alpha = .05

Tests of Between-Subjects Effects

Dependent Variable: QoLS_Phys

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1.388[a]	3	.463	2.754	.043
Intercept	3138.574	1	3138.574	18685.381	.000
CONDITION	.902	1	.902	5.367	.021
TIME	.142	1	.142	.848	.358
CONDITION * TIME	.347	1	.347	2.066	.152
Error	46.024	274	.168		
Total	3185.383	278			
Corrected Total	47.411	277			

a. R Squared = .029 (Adjusted R Squared = .019)

b. Computed using alpha = .05

Tests of Between-Subjects Effects

Dependent Variable: QoLS_SocPsych

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.101[a]	3	.034	.224	.880
Intercept	1144.091	1	1144.091	7585.116	.000
CONDITION	.027	1	.027	.179	.672
TIME	.028	1	.028	.184	.669
CONDITION * TIME	.046	1	.046	.306	.581
Error	41.328	274	.151		
Total	1185.660	278			
Corrected Total	41.430	277			

a. R Squared = .002 (Adjusted R Squared = -.008)

b. Computed using alpha = .05

Tests of Between-Subjects Effects

Dependent Variable: QoLS_PosAtt

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.706[a]	3	.235	1.360	.255
Intercept	3270.962	1	3270.962	18904.339	.000
CONDITION	.216	1	.216	1.248	.265
TIME	.114	1	.114	.659	.418
CONDITION * TIME	.379	1	..379	2.189	.140
Error	47.409	275274	.173		

Total	3318.864	278			
Corrected Total	48.115	277			

a. R Squared = .015 (Adjusted R Squared = .004)

b. Computed using alpha = .05

Topic guide

Teacher Interviews

Topic guide.

Pre interview:

- Explain that the interview is confidential and that they will not be identified in any way. So they can be completely honest.
- Explain that they can stop at any time, should they want to.
- Ask if they have any questions,
- Have they signed the consent form?
- We are only focusing on what is new in the school.

Did you know about SC before?

What elements did you focus on or were new to the classroom practice?

- Growth mind-set: Do the Pupils understand and can explain this. Using the muscles to forward there learning.
- Learning outcomes: Out of context learning should be always seen
- Success criteria: Co constructed and visible to the class. Examples
- Talk partners: Random selection, weekly change.
- Differentiated challenges/ mixed ability grouping: For all pupils, for all lessons
- Visualizer: Pit stops, uses examples of work, peer feedback
- Successful starts to lessons, prior knowledge questions.
- Involving pupils in the planning: Independence within their own learning
- Elimination of rewards: barriers or facilitators: Praise effort, Celebrate mistakes
- Feedback: Have they change the way in which they give feedback, has it reduced the workload

Effect in the class/school both academic and behavioural Social Effects
Personal/professional development and attitudes
Challenges to this way of working
Dissemination in the school
What do you think of the peer mentoring model of delivery?

How much time did it take you to implements as in extra time on top of your normal prevision? Did it save time e.g. more effective marking.

What extra materials were needed to deliver the SC?

Is it now just part of the way that you work?

Have you had any feedback from parents?

Can you see that some elements work with the new curriculum?

Appendix F

Observation checklist

Formative Assessment Classroom Guide
What part of the lesson are you observing?

First 15 minutes (F)		Middle (M)		Last 15 minutes (L)	

Depending on the answer to the above question observers will be looking for different strategies/techniques and therefore only the relevant strategies/techniques should show.

		F	M	L
1	Is it clear what the teacher intends the students to learn? (e.g. sharing learning outcomes with pupils)	/	/	/
2	Does the teacher identify student learning needs? (e.g., eliciting prior knowledge)	/		
3	Do students understand what criteria will make their work successful? (e.g., success criteria, sharing learning objectives, co- constructing success criteria)	/	/	/
4	Are students chosen at random to answer questions?			
5	Does the teacher ask questions that make students think? (e.g., talk partners)	/	/	/
6	Does the teacher give students time to think after asking a question? (e.g., talk partners)	/	/	/
7	Does the teacher allow time for students to elaborate their responses?		/	/
8	Is a whole-class response system used? (e.g., thumbs up strategy)			
9	Is teaching adjusted after gathering pupil feedback (data collection)? (e.g., teacher using immediate pupil feedback to adjust the lesson)		/	/
10	Is there more student talk than teacher talk?		/	/
11	Are most students involved in answering questions?	/	/	/
12	Are students supporting each other's learning? (e.g., talk partners , peer assessment, sharing work under the visualizer)	/	/	/
13	Is there evidence that various forms of teacher feedback advance student learning? (e.g., evidence of pupils editing classwork for improvement, using teacher feedback)	/	/	/

14	Do students take responsibility for their own learning? (e.g. Choosing tasks according to differentiated challenges, Self-assessment, and improvement, planning their own way of working / problem solving , working on or creating new personal targets)	/	/	/
15	Does the teacher give oral formative feedback?	/	/	/
16	Does the teacher find out what the students have learned before they leave the room?(e.g., talk partners activities , notes from pupils exit messages, questioning)			/

There are 5 possible responses to each of the questions. All questions should be answered. There is only one possible answer to each question.

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented

1. Is it clear what the teacher intends the students to learn?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
When the teacher deliberately defers sharing learning intentions until later	Teacher makes no attempt to communicate learning intentions to students	Teacher states or writes learning intentions but does not check that they make sense to students	Learning intentions clear and communicated to at least half the students	Learning intentions clear and successfully communicated to almost all students

17. Does the teacher identify student learning needs?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Only when the lesson starts with an activity	Evidence of students' prior knowledge would be useful but not elicited	Elicits relevant evidence from only a few students in order to decide where to start the lesson	Refers back to what students know and still need to learn from last period or at start of lesson	Refers back to what students still need to learn and builds on this at the start of the lesson

18. Do students understand what criteria will make their work successful?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
It is always applicable for students to know what makes their work successful	No students appear to know what will make their work successful	Some students are clear about the success criteria for the lesson	Most students are clear about the success criteria for the lesson	Almost all students are clear about the success criteria for the lesson

19. Are students chosen at random to answer questions?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
The teacher asks only whole-class response questions	Students put hands up and teacher chooses one of them to answer a question	Teacher chooses students to answer questions, whether they have their hands up or not	Teacher sometimes uses a randomizing method to choose students to answer questions	Teacher usually uses a randomizing method to choose students to answer questions

20. Does the teacher ask questions that make students think?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Teacher asks no questions	Only recall questions asked	At least one question makes students think	Some questions make students think	Students give thoughtful answers to all questions

21. Does the teacher give students time to think after asking a question?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
When the teacher asks only questions that do not require thinking	Teacher passes question to another student, or expands the question if the student does not answer straight away	Teacher occasionally waits for 3 seconds for a student to think before answering a question	Teacher usually waits at least 3 seconds before calling on a student to answer a question	Teacher always waits at least 3 seconds before calling on a student to answer a question

22. Does the teacher allow time for students to elaborate their responses?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
When the teacher asks only questions that do not require elaboration	The teacher evaluates or moves on before, or as soon as, student finishes	Teacher occasionally waits at least 3 seconds after student answers a question to allow for elaboration	Teacher usually waits at least 3 seconds after student answers a question to allow for elaboration	Teacher always waits at least 3 seconds after student answers a question to allow for elaboration

23. Is a whole-class response system used?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
This is not a suitable technique for this part of the lesson	Whole class response system not used when it would have been a suitable technique	At least one question asked where all students show their answer, but not all students participate	Some questions asked where all students show their answer, eg using fingers, ABCD cards, white boards	Many questions asked where all students show their answer, eg using fingers, ABCD cards, white boards

24. Is teaching adjusted after gathering feedback from pupils (data collection)?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
It is appropriate that no data collection is made during this part of the lesson	Teacher makes no adjustments, even though it would have been appropriate	Data collected but not obviously used	Data collected and appears to be used to adjust teaching to some degree	Teacher comments to class on how data helps her decide on next phase of lesson

25. Is there more student talk than teacher talk?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Only when teacher talk is appropriate for all of this part of the lesson	Teacher talk dominates this part of the lesson, when student talk would have been useful	Students talk as much as the teacher	Students talk more than the teacher	Purposeful student talk dominates the time

26. Are most students involved in answering questions?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented

No questions asked during the observation	Questions would have been appropriate but not asked	Only a few students involved in answering questions	About half of the class involved in answering questions	Most students involved in answering questions

27. Are students supporting each other's learning?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Only quiet individual work observed	Would have been great to see students supporting each other!	Students observed attempting to support other students, but inappropriately	Some students supporting each other effectively	Many students supporting peers' leaning effectively

28. Is there evidence that various forms of teacher feedback advance student learning?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Only at start of semester with a new teacher	No teacher feedback	Some teacher feedback, but not helpful	At least one example of feedback used to advance learning	Many examples of teacher feedback that advances student learning

29. Do students take responsibility for their own learning?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
It is always applicable for students to be taking responsibility for their own learning	No students are working independently, and no students seek help	Few students are working independently, or seek help at appropriate times	Many students are working independently, or seek help at appropriate times	Almost all students are effectively working independently, or seek help at appropriate times

30. Does the teacher provide oral formative feedback?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
Providing oral feedback would not be appropriate	Teacher provides no oral feedback to students despite it being appropriate	Teacher provides oral feedback to students that is unlikely to help them progress	Teacher provides helpful oral feedback to some students	Teacher provides helpful oral feedback to many students

31. Does the teacher find out what the students have learned before they leave the room?

Not applicable	Applicable but not observed	Observed but poorly implemented	Observed and reasonably implemented	Observed and well implemented
It is always applicable to find out what students have learned	Students leave room without any feedback to the teacher on what they have learned	Self-reporting by students with no check on the validity of responses	Teacher knows how much some of the students have learned at the end of the lesson	Teacher and students leave the room clear about what has been learned

Appendix G Calculations for BAU

Primary Schools				
	Anglesey	Conwy	Wrexham	
Number of Schools	40	62	67	
Average teacher wage	34034.44	34034.44	34034.44	nasuwt
Salary increase	2%	2%	2%	ASCL
Pension and NI contributions	28%	28%	28%	
Spending on other	16.36%	Does not state take average between the other counties	20.01%	For Wexham they do not have the breakdown
Maintenance fees	9.49%			This is the amount the council calculate that is needed for school (its part of their funding formula)
Total budget	13,369,000	22,377,000	45,272,030	Does not include SES or ISB
Pupil teacher Ratio (PTR)	18.5	17.7	19.2	
Number of pupils in county	5562	8486	12,343	Council budget document
Number of teachers	199	327	470	Primary only
Average eFSM	15.5	17.5	18	Not needed as comes from different funding area.
	34034.44 28% = 9530	34034.44 28% = 9530	34034.44 28% = 9530	
	34034.44 2% = 681	34034.44 2% = 681	34034.44 2% = 681	
Total salary	44,245	44,245	44,245	

School running costs	25.85	25.85 + 20.01 / 2 = 22.93	20.01	
Total school running costs	13,369,000 25.85 % = 3,360,500	22,377,000 22.93 = 5,044,6000	45,272,030 20.01 % = 9,004,500	
Total	3,360,500	5,044,600	9,004,500	
Total cost per teacher running costs	3,360,500/199 = 16,887	5,044,600 / 327 = 15,426	9,004,500/470 = 19,159	
Total	16,887	15,426	19,159	
Total salary with running costs	44245 + 16,887 = 61134	44245 + 15426 = 59671	44245 + 19,159 = 63,404	
Total	61,134	59,671	63,404	
Including a head teacher in the calculation				
	199 + 40 = 239	327 + 62 = 389	470 + 67 = 537	
	3,360,500/ 239 = 14,061	5,044,600 / 389 = 12,968	9,004500 / 537 = 16,768	
Total salary with running cost and assumption that each school has a head teacher	44245 + 14061 = 58,306	44245 + 12,968 = 56,313	44245 + 16,768 = 61,013	
Total	58,306	56,313	61,013	

Per pupil	58,306/18.5 = 3152	56313/17.7 = 3218	61013/19.2 = 3178	
NASUWT suggest that teacher can only be asked to teach a maximum of 1265 hours over 195 days	58,306/1265 = 46	56313/1265 = 45	61013/1265 = 48	
Cost for a teacher per hour	£46	£45	£48	
Average across all three LAs	**Cost 2018- 2019**	**Cost 2020- 2021**		
Teacher salary	61403	63924		
Teacher salary with HT	58,544	60947		
Cost per pupil (HT inc)	3183	3314		
Cost per hour	46	47.89		